DOCUMENT RESUME

ED 459 856 IR 058 408

AUTHOR Covey, Denise Troll
TITLE Usage and Usability Assessment: Library Practices and Concerns.
INSTITUTION Council on Library and Information Resources, Washington, DC. Digital Library Federation.
ISBN ISBN-1-887334-89-0
PUB DATE 2002-01-00
NOTE 100p.
AVAILABLE FROM Council on Library and Information Resources, Digital Library Federation, 1755 Massachusetts Ave., NW, Suite 500, Washington, DC 20036 ($20). Tel: 202-939-4750; Fax: 202-939-4765; e-mail: info@clir.org; Web site: http://www.clir.org.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Electronic Libraries; *Evaluation; *Evaluation Methods; Evaluation Research; Evaluation Utilization; Focus Groups; Information Services; *Information Utilization; Research Methodology; Use Studies; Users (Information)
IDENTIFIERS Electronic Resources; *Transaction Log Analysis; *Usability

ABSTRACT

This report presents the results of a survey of Digital Library Federation members that examined the kinds of assessments they were conducting, what they did with the results, and what worked well or not so well. The first section of the report is an introduction that describes the report structure and summarizes challenges in assessment. The second section presents results related to user studies, including surveys (questionnaires), focus groups, user protocols, and other effective research methods. The third section addresses the following questions related to usage studies of electronic resources: (1) What is transaction log analysis? (2) Why do libraries conduct transaction log analysis? (3) How do libraries conduct transaction log analysis? (4) Who uses the results of transaction log analysis? and (5) What are the issues, problems, and challenges with transaction log analysis? The fourth section covers general issues and challenges in planning and implementing a research project. The fifth section presents conclusions and considers future directions. Appendices include: references and a selected bibliography; a list of participating institutions; the survey questions; and traditional input, output, and outcome measures. (MES)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 459 856

Usage and Usability Assessment: Library Practices and Concerns

by Denise Troll Covey

January 2002

Digital Library Federation
Council on Library and Information Resources
Washington, D.C.

IR058408

BEST COPY AVAILABLE

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

B.H. Leney

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

About the Author

Denise Troll Covey is associate university librarian of arts, archives, and technology at Carnegie Mellon University. In 2000-2001 she was also a distinguished fellow in the Digital Library Federation, leading the initiative on usage, usability, and user support. Her professional work focuses on the research and development of digital library collections, services, and software; and assessment practices, copyright permissions, and change management as they relate to digital libraries. Covey has academic degrees in theology, philosophy, and rhetoric. Her graduate work emphasized the history of information storage and retrieval.

Acknowledgments

The author and the Digital Library Federation sincerely thank the 71 individuals who participated in the DLF telephone survey. Their time, experiences, concerns, and questions about library use and usability made this report possible. If the report facilitates discussion and research and encourages the development of benchmarks or best practices, it is because so many talented people shared their rewards and frustrations in trying to understand what is happening in their libraries and how to serve their users better.

ISBN 1-887334-89-0

Published by:

Digital Library Federation
Council on Library and Information Resources
1755 Massachusetts Avenue, NW, Suite 500
Washington, DC 20036

Web site at http://www.clir.org

Additional copies are available for $20 per copy. Orders must be placed online through CLIR's Web site.

The paper in this publication meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials ANSI Z39.48-1984.

Copyright 2002 by the Council on Library and Information Resources. No part of this publication may be reproduced or transcribed in any form without permission of the publisher. Requests for reproduction should be submitted to the Director of Communications at the Council on Library and Information Resources.

Contents

Preface v

1. Introduction 1
    1.1. Report Structure 1
    1.2. Summary of Challenges in Assessment 2

2. User Studies 5
    2.1. Surveys (Questionnaires) 6
        2.1.1. What Is a Survey Questionnaire? 6
        2.1.2. Why Do Libraries Conduct Surveys? 7
        2.1.3. How Do Libraries Conduct Surveys? 8
        2.1.4. Who Uses Survey Results? How Are They Used? 10
        2.1.5. What Are the Issues, Problems, and Challenges With Surveys? 12
            2.1.5.1. The Costs and Benefits of Different Types of Surveys 12
            2.1.5.2. The Frequency of Surveys 12
            2.1.5.3. Composing Survey Questions 12
            2.1.5.4. Lack of Analysis or Application 14
            2.1.5.5. Lack of Resources or Comprehensive Plans 14

    2.2. Focus Groups 15
        2.2.1. What Is a Focus Group? 15
        2.2.2. Why Do Libraries Conduct Focus Groups? 17
        2.2.3. How Do Libraries Conduct Focus Groups? 18
        2.2.4. Who Uses Focus Group Results? How Are They Used? 21
        2.2.5. What Are the Issues, Problems, and Challenges With Focus Groups? 22
            2.2.5.1. Unskilled Moderators and Observers 22
            2.2.5.2. Interpreting and Using the Data 22

    2.3. User Protocols 23
        2.3.1. What Is a User Protocol? 23
        2.3.2. Why Do Libraries Conduct User Protocols? 24
        2.3.3. How Do Libraries Conduct User Protocols? 25
        2.3.4. Who Uses Protocol Results? How Are They Used? 28
        2.3.5. What Are the Issues, Problems, and Challenges With User Protocols? 29
            2.3.5.1. Librarian Assumptions and Preferences 29
            2.3.5.2. Lack of Resources and Commitment 29
            2.3.5.3. Interpreting and Using the Data 29
            2.3.5.4. Recruiting Participants Who Can Think Aloud 30

2.4. Other Effective Research Methods 30
2.4.1. Discount Usability Research Methods 30
2.4.1.1. Heuristic Evaluations 30
2.4.1.2. Paper Prototypes and Scenarios 32
2.4.2. Card-Sorting Tests 32

3. Usage Studies of Electronic Resources 33
3.1. What Is Transaction Log Analysis? 33

3.2. Why Do Libraries Conduct Transaction Log Analysis? 35

3.3. How Do Libraries Conduct Transaction Log Analysis? 35
3.3.1. Web Sites and Local Digital Collections 35
3.3.2. OPAC and Integrated Library Systems 37

3.4. Who Uses the Results of Transaction Log Analysis?
How Are They Used? 37
3.4.1. Web Sites and Local Digital Collections 37
3.4.2. OPAC and Integrated Library Systems 39
3.4.3. Remote Electronic Resources 40

3.5. What Are the Issues, Problems, and Challenges With Transaction
Log Analysis? 40
3.5.1. Getting the Right (Comparable) Data and Definitions 40
3.5.1.1. Web Sites and Local Digital Collections 40
3.5.1.2. OPAC and Integrated Library Systems 41
3.5.1.3. Remote Electronic Resources 42
3.5.2. Analyzing and Interpreting the Data 43
3.5.3. Managing, Presenting, and Using the Data 43

4. General Issues and Challenges 44
4.1. Issues in Planning a Research Project 44

4.2. Issues in Implementing a Research Project 48
4.2.1. Issues in Sampling and Recruiting Research Subjects 48
4.2.2. Issues in Getting Approval and Preserving Anonymity 51

5. Conclusions and Future Directions 53

APPENDIX A: References and Selected Bibliography 61

APPENDIX B: Participating Institutions 75

APPENDIX C: Survey Questions 76

APPENDIX D: Traditional Input, Output, and Outcome Measures 77

Preface

Making library services available online is not only expensive; it is also very risky. The library's roles there are not at all clear. Neither are its relationships with users or with other information services. There is little information about how library users behave in a network environment, how they react to online library services, and how they combine those services with others such as search engines like Google, bookstores like Amazon, Internet gateways like Voice of the Shuttle, and instructional technologies like WebCT or Blackboard. Digital libraries are still relatively immature—most are still at a stage where limited experimentation is more important than well-informed strategic planning. While libraries have excelled at assessing the development and use of their traditional collections and services, comparable assessments of online collections and services are more complicated and less well understood.

Against this backdrop, the Digital Library Federation (DLF) has committed to driving forward a research process that will provide the information that libraries need to inform their development in a networked era. The goals of this process are:

- to develop a better understanding of methods effective in assessing use and usability of online scholarly information resources and information services; and
- to create a baseline understanding of users' needs to support strategic planning in an increasingly competitive environment for academic libraries and their parent institutions.

This report is an initial step in achieving the first of these goals. It offers a survey of the methods that are being deployed at leading digital libraries to assess the use and usability of their online collections and services. Focusing on 24 DLF member libraries, the study's author, Distinguished DLF Fellow Denise Troll Covey, conducted numerous interviews with library professionals who are engaged in assessment. In these interviews, Covey sought to document the following:

- why digital libraries assessed the use and usability of their online collections and services
- what aspects of those collections and services they were most interested in assessing
- what methods the libraries used to conduct their assessments
- which methods worked well and which worked poorly in particular kinds of assessments
- how assessment data were used by the library, and to what end
- what challenges libraries faced in conducting effective assessments

The result is a report on the application, strengths, and weaknesses of assessment techniques that include surveys, focus groups, user protocols, and transaction log analysis. Covey's work is also an essential methodological guidebook. For each method that she covers, she is careful to supply a definition, explain why and how libraries use the method, what they do with the results, and what problems they encounter. The report includes an extensive bibliography on more detailed methodological information, and descriptions of assessment instruments that have proved particularly effective. Examples are available on the Web for all to see, and potentially to modify and use. The work concludes with a review of the challenges that libraries face as they seek to gather and use reliable information about how their online presence is felt. These concluding remarks will be of general interest and are recommended to senior library managers as well as to those more directly involved with assessment activities.

Given its practical orientation, *Usage and Usability* is an ideal launching pad for CLIR's new series, *Tools for Practitioners*. The series emphasizes the immediate, the practical, and the methodological. As it develops, it will include work that, like Covey's, appeals to and provides guidance for particular professional audiences.

Daniel Greenstein
Director, Digital Library Federation

1. INTRODUCTION

As the needs and expectations of library users change in the digital environment, libraries are trying to find the best ways to define their user communities, understand what they value, and evolve digital library collections and services to meet their demands. In part, this effort requires a closer, more formal look at how library patrons use and respond to online collections and services.

To synthesize and learn from the experiences of leading digital libraries in assessing use and usability of online collections and services, the Digital Library Federation (DLF) undertook a survey of its members. From November 2000 through February 2001, the author conducted interviews with 71 individuals at 24 of the 26 DLF member institutions (representing an 86 percent response rate at the 24 institutions). Participants were asked a standard set of open-ended questions about the kinds of assessments they were conducting; what they did with the results; and what worked well or not so well. Follow-up questions varied, based on the work being done at the institution; in effect, the interviews tracked the efforts and experiences of those being interviewed.

The results of the survey reveal the assessment practices and concerns of leading digital libraries. They are not representative of all library efforts; however, they do show trends that are likely to inform library practice. The study offers a qualitative, rather than quantitative, assessment of issues and practices in usage and usability data gathering, analysis, interpretation, and application.

1.1. Report Structure

The survey indicates significant challenges to assessing use and usability of digital collections and services. The rest of Section 1 summarizes these challenges. Subsequent sections elaborate on these challenges and draw on examples from the assessment efforts of DLF libraries. Sections 2 and 3 describe libraries' experiences using popu-

lar methods to conduct user studies, such as surveys, focus groups, user protocols, and transaction log analysis. The report explains what each of these methods entails, its advantages and disadvantages, why and how libraries use it, the problems encountered, and the lessons libraries have learned from experience. Section 4 covers general issues and challenges in conducting research, including sampling and recruiting representative research subjects, getting Institutional Review Board (IRB) approval to conduct research with human subjects, and preserving user privacy. Section 5 summarizes the conclusions of the study and suggests an agenda for future discussion and research. Appendix A provides a selected bibliography. A list of institutions participating in the survey appears in Appendix B, while Appendix C lists the interview questions. An overview of more traditional library input, output, and outcome assessment efforts, and the impact of digital libraries on these efforts, is provided in Appendix D; this information is designed to help the reader position the information in this report within the context of library assessment practices more generally.

To preserve the anonymity of DLF survey respondents and respect the sensitivity of the research findings, the report does not associate institution names with particular research projects, incidents, or results. The word "faculty" is used to refer to teachers and professors of for-credit academic courses. The word "librarian" is used, regardless of whether librarians have faculty status in their institutions, or, indeed, whether they hold an MLS degree.

1.2. Summary of Challenges in Assessment

DLF respondents shared the following concerns about the efficiency and efficacy of their assessment efforts:

- Focusing efforts to collect only meaningful, purposeful data
- Developing the skills to gather, analyze, interpret, present, and use data
- Developing comprehensive assessment plans
- Organizing assessment as a core activity
- Compiling and managing assessment data
- Acquiring sufficient information about the environment to understand trends in library use

Collecting only meaningful, purposeful data. Libraries are struggling to find the right measures on which to base their decisions. DLF respondents expressed concern that data are being gathered for historical reasons or because they are easy to gather, rather than because they serve useful, articulated purposes. They questioned whether the sheer volume of data being gathered prohibits their careful analysis and whether data are being used to their full advantage. Working with data is essential, time-consuming, and costly–so costly that libraries are beginning to question, and in some cases even measure, the costs and benefits of gathering and analyzing different data. Respondents know that they need new measures and composite measures to capture the extent of their activities in

both the digital and traditional realms. Adding new measures is prompting many DLF sites to review their data-gathering practices. The libraries are considering, beginning, or completing needs assessments of the data they currently gather, or think they should gather, for internal and external purposes. If such information is not needed for national surveys or not useful for strategic purposes, chances are it will no longer be gathered, or at least not gathered routinely. However, deciding what data should be gathered is fraught with difficulties. Trying to define and measure use of services and collections that are rapidly changing is a challenge. The fact that assessment methods evolve at a much slower rate than do the activities or processes they are intended to assess compounds the problem. How can libraries measure what they do, how much they do, or how well they do, when the boundaries keep changing?

Developing skills to gather, analyze, interpret, present, and use data. Several DLF respondents commented that they spend a great deal of time gathering data but do not have the time or talent to do anything with this information. Even if libraries gather the right measures for their purposes, developing the requisite skills to analyze, interpret, present, and use the data are separate challenges. For example, how do you intelligibly present monthly usage reports on 8,000 electronic journals? The answer is you don't. Instead, you present the statistics on the top 10 journals, even though this severely limits the dissemination and application of data painstakingly gathered and compiled. Though DLF respondents indicated that they are learning slowly from experience how to make each research method work better for their purposes, many said they need methodological guidance. They need to know *what* sampling and research methods are available to recruit research subjects and assess use and usability of the digital library, *which* methods are best suited for which purposes, and *how* to analyze, interpret, present, and use the quantitative and qualitative data they gather to make effective decisions and strategic plans.

Developing comprehensive assessment plans. Planning assessment from conception through follow-up also presents challenges. Ideally, the research process should flow seamlessly—from deciding to gather data to developing and implementing plans to use the data. In reality, however, DLF respondents reported frequent breakdowns in this process. Breakdowns occur for a number of reasons. It may be that something went awry in the planning or scheduling of the study. People assigned responsibility for certain steps in the process may lack the requisite skills. Staff turnover or competing priorities may intervene. Respondents also made it clear that the more people involved in the research process, the longer it takes. The longer the process takes, the more likely it is that the results will be out of date, momentum will be lost, or other phenomena will intrude before the results are implemented. Finally, if the study findings go unused, there will be less enthusiasm for the next study, and participation is likely to decrease. This applies both to the people conducting the study and to the research subjects. Conducting a study creates expec-

tations that something will be done with the results. When the results are not applied, morale takes a hit and human and financial resources are wasted. Participants lose confidence, and the study planners lose credibility.

Organizing assessment as a core activity. DLF respondents well understood that in an environment of rapid change and limited resources, libraries cannot afford these outcomes from their assessment efforts. They also seemed to understand that the way in which an assessment is organized affects the outcome. At some institutions, user studies are centralized and performed by recently hired experts in the field. At others, user studies are decentralized and performed systemwide; they involve efforts to teach librarians and staff throughout the organization how to conduct research using different methods. Still other institutions, sparked by the interests of different personnel, take an ad hoc approach to user studies. A few libraries have established usability testing programs and laboratories. If the goal is a culture of assessment, then making assessment a core activity and allocating human and financial resources to it is essential. The key is not *how* a study is organized, but *that* it is organized and supported by commitment from administrators and librarians. Comments from DLF respondents suggested that given sufficient human and financial resources, requisite skills could be acquired, guidelines and best practices developed, and assessments conducted routinely, efficiently, and effectively enough to keep pace with the pace of change.

Compiling and managing assessment data. Many DLF respondents expressed concern about the effort required to compile and manage data collected by different people and assessments. Libraries need a simple way to record and analyze quantitative and qualitative data and to generate statistical reports and trend lines. Several DLF sites have developed or are developing a management information system (MIS) to compile and manage statistical data. They are wrestling with questions about how long data should be kept, how data should be archived, and whether one system can or should manage data from different kinds of assessments. Existing systems typically have a limited scope. For example, one site has a homegrown desktop reporting tool that enables library staff to generate ad hoc reports from data extracted and to update them regularly from the integrated library system. Users can query the data and run cross-tabulations. The tool is used for a variety of purposes, including analysis of collection development, materials expenditures, and the productivity of the cataloging department. Reports can be printed, saved, or imported into spreadsheets or other applications for further analysis or manipulation. New systems being developed appear to be more comprehensive; for example, they attempt to assemble statistical data from all library departments. The ability to conduct cross-tabulations of data from different departments and easily generate graphics and multiyear trend lines are important features of the new systems.

Acquiring sufficient information about the environment to understand trends in library use. Several DLF respondents noted that

emerging new measures will assess *how* library use is changing in the networked environment, but these measures will not explain *why* library use is changing. Academic libraries need to know how students and faculty find information, what resources they use that the libraries do not provide, why they use these resources, and what they do with the information after they find it. This knowledge would provide a context for interpreting existing data on shifting patterns of library use and facilitate the development of collections, services, and tools that better meet user needs and expectations. Library user studies naturally focus on the use and usability of library collections, services, and Web sites. The larger environment remains unexplored.

2. USER STUDIES

DLF respondents devoted the bulk of their discussion to user studies, reflecting the user-centered focus of their operations. One respondent referred to the results of user studies as "outcome" measures because, although they do not measure the impact of library use on student learning or faculty research, they do indicate the impact of library services, collections, facilities, and staff on user experiences and perceptions.

Libraries participating in the DLF survey organize, staff, and conduct user studies differently. Some take an ad hoc approach; others use a more systematic approach. Some sites have dedicated staff experts in research methodologies who conduct user studies; others train staff throughout the libraries to conduct user studies. Some libraries take both approaches. Some have consulted experts on their campuses or contracted with commercial firms to develop research instruments and analyze the results. For example, libraries participating in the DLF survey have recruited students in library science and human-computer interaction to conduct user studies or hired companies such as Websurveyor.com or Zoomerang.com to host Web-based surveys and analyze the data. Libraries that conduct user studies use spreadsheet, database, or statistical analysis software to manage and analyze the data. In the absence of standard instruments, guidelines, or best practices, institutions either adapt published efforts to local circumstances or make their own. There is clearly a flurry of activity, some of it not well organized or effective, for various reasons discussed elsewhere in this report.

Learning how to prepare research instruments, analyze and interpret the data, and use the results is a slow process. Unfortunately, however, the ability to quickly apply research results is often essential, because the environment changes quickly and results go out of date. Many DLF respondents reported instances where data languished without being analyzed or applied. They strongly cautioned against conducting research when resources and interest are insufficient to support use of the results. Nevertheless, DLF libraries are conducting many user studies employing a variety of research methods. The results of these studies run the gamut: they may reinforce

librarian understanding of what users need, like, or expect; challenge librarian assumptions about what people want; or provide conflicting, ambiguous, misleading, or incomplete information that requires follow-up research to resolve or interpret. Multiple research methods may be required to understand fully and corroborate research results. This exacerbates an already complicated situation and can frustrate staff. Resources may not be available to conduct follow-up studies immediately. In other cases, new priorities emerge that make the initial study results no longer applicable; in such a case, any attempt at follow-up is worthless. Moreover, even when research data have been swiftly analyzed, interpreting the results and deciding how to apply them may be slowed if many people are involved in the process or if the results challenge long-held assumptions and preferences of librarians. Finally, even when a plan to use the results is in hand, implementation may pose a stumbling block. The longer the entire research process takes, from conception to implementing the results, the more likely the loss of momentum and conflict with other priorities, and the greater the risk that the process will break down and the effort will be wasted. The issue appears to be related to the internal organization and support for the library's assessment effort.

To help libraries understand and address these concerns, this section of the report describes popular user study methods, when and why DLF libraries have used them, where they succeeded, and where they failed. Unless otherwise noted, all claims and examples derive from the DLF interviews. The focus is surveys, focus groups, and user protocols, which are the methods DLF libraries use most often. Heuristic evaluations, paper prototypes and scenarios, and card-sorting exercises are also described because several DLF institutions have also used these methods successfully.[1]

2.1. Surveys (Questionnaires)

2.1.1. What Is a Survey Questionnaire?

Survey questionnaires are self-administered interviews in which the instructions and questions are sufficiently complete and intelligible for respondents to act as their own interviewers.[2] The questions are simply stated and carefully articulated to accomplish the purpose for which the survey is being conducted. Survey questions typically force respondents to choose from among alternative answers provided or to rank or rate items provided. Such questions enable a simple quantitative analysis of the responses. Surveys can also ask open-ended questions to gather qualitative comments from the respondents.

[1] To give the reader a better understanding of the care with which user studies must be designed and conducted, sample research instruments may be viewed at www.clir.org/pubs/reports/pub105/instr.pdf.

[2] Much of the information in this section is taken from Chadwick, Bahr, and Albrecht 1984.

Surveys are an effective way to gather information about respondents' previous or current behaviors, attitudes, beliefs, and feelings. They are the preferred method to gather information about sensitive topics because respondents are less likely to try to please the researcher or to feel pressured to provide socially acceptable responses than they would in a face-to-face interview. Surveys are an effective method to identify problem areas and, if repeated over time, to identify trends. Surveys cannot, however, establish cause-effect relationships; and the information they gather reveals little if anything about contextual factors affecting the respondents. Additional research is usually required to gather the information needed to determine how to solve the problems identified in a survey.

The primary advantage of survey questionnaires is economy. Surveys enable researchers to collect data from large numbers of respondents in relatively short periods of time at relatively low cost. Surveys also give respondents time to think about the questions before answering and often do not require respondents to complete the survey in one sitting.

The primary disadvantage of survey questionnaires is that they must be simple, impersonal, and relatively brief. If the survey is too long or complex, respondents may get tired and hurriedly answer or skip questions. The response rate and the quality of responses decline if a survey exceeds 11 pages (Dillman 1978). Instructions and questions must be carefully worded in language meaningful to the respondents, because no interviewer is present to clarify the questions or probe respondents for additional information. Finally, it is possible that someone other than the selected respondent may complete the survey. This can skew the results from carefully selected samples. (For more about sampling, see section 4.2.1.) When necessary, survey instructions may explicitly ask that no one complete the survey other than the person for whom it is intended.

2.1.2. Why Do Libraries Conduct Surveys?

Most of the DLF respondents reported conducting surveys, primarily to identify trends, "take the temperature" of what was happening among their constituencies, or get a sense of their users' perceptions of library resources. Occasionally they conduct surveys to compare themselves with their peers. In summary, DLF libraries have conducted surveys to assess the following:

- Patterns, frequency, ease, and success of use
- User needs, expectations, perspectives, priorities, and preferences for library collections, services, and systems
- User satisfaction with vendor products, library collections, services, staff, and Web sites
- Service quality
- Shifts in user attitude and opinion
- Relevance of collections or services to the curriculum

A few respondents reported conducting surveys as a way to market their collections and services; others commented that this

was an inappropriate use of survey research. One respondent referred to this type of survey as "push polling" and stated that there were easier, more appropriate ways than this to market what the library offers.

The data gathered from surveys are used to inform decision making and strategic planning related to the allocation of financial and human resources and to the organization of library units. Survey data also serve political purposes. They are used in presentations to faculty senates, deans' councils, and library advisory boards as a means to bolster support for changes in library practice. They are also used in grant proposals and other requests for funding.

2.1.3. How Do Libraries Conduct Surveys?

DLF respondents reported that they conduct some surveys routinely; these include annual surveys of general library use and user priorities and satisfaction. Other surveys are conducted sporadically; in this category might be, for example, a survey to determine user satisfaction with laptop-lending programs. The library administrator's approval is generally required for larger, more formal, and routine surveys. Smaller, sporadic, less expensive surveys are conducted at the discretion of middle managers.

Once the decision has been made to conduct a survey, libraries convene a small group of librarians or staff to prepare the survey instructions and questionnaire, determine the format of the survey (for example, print, e-mail, Web-based), choose the sampling method, identify the demographic groups appropriate for the research purpose, determine how many participants to recruit in each group and decide how to recruit them, and plan the budget and timetable for gathering, analyzing, interpreting, and applying the data. A few DLF respondents reported using screening questionnaires to find experienced or inexperienced users, depending on the purpose of the study.

Different procedures are followed for formal surveys than for small surveys. The former require more work. Because few libraries employ survey experts, a group preparing a formal survey might consult with survey experts on campus to ensure that the questions it has drafted will gather the information needed. The group might consult with a statistician on campus to ensure that it recruits enough participants to gather statistically significant results. When a survey is deemed to be extremely important and financial resources are available, an external consulting or research firm might be hired. Alternatively, libraries with adequate budgets and sufficient interest in assessment have begun to use commercial firms such as Websurveyor.com to conduct some surveys.

If the survey is to be conducted in-house, time and financial constraints and the skills of library staff influence the choice of survey format. Paper surveys are slow and expensive to conduct. Follow-up may be needed to ensure an adequate response rate. Respondents are not required to complete them in one sitting; for this reason, paper surveys may be longer than electronic surveys. E-mail surveys are

less expensive than paper surveys; otherwise, their advantages are similar. Web-based surveys might be the least expensive to conduct, particularly if scripts are available to analyze the results automatically. They also offer several other advantages. For example, they can be edited up to the last minute, and the capabilities of the Web enable sophisticated branching and multimedia surveys, which are difficult or even impossible, in other formats. Both Web and e-mail surveys are easier to ignore than are paper surveys, and they assume participants have computer access. Web surveys have the further disadvantage that they must be completed in one sitting, which means they must be relatively short. They also require HTML skills to prepare and, if results are to be analyzed automatically, programming skills. Whether Web-based surveys increase response rate is not known. One DLF library reported conducting a survey in both e-mail and Web formats. An equal number of respondents chose to complete the survey in each format.

Considerable time and effort should be spent on preparing the content and presentation of surveys. Instructions and questions must be carefully and unambiguously worded and presented in a layout that is easy to read. If not, results will be inaccurate or difficult or impossible to interpret, worse yet, participants may not complete the survey. The choice of format affects the amount of control libraries have over the presentation or appearance of the survey. Print offers the most control; with e-mail and Web-based formats, there is no way for the library to know exactly what the survey will look like when it is viewed using different e-mail programs or Web browsers. The group preparing e-mail or Web surveys might find it helpful to view the survey using e-mail programs and Web browsers available on campus to ensure that the presentation is attractive and intelligible.

Libraries pilot test survey instructions and questions with a few users and revise them on the basis of test results to solve problems with vocabulary, wording, and the layout or sequence of the questions. Pilot tests also indicate the length of time required to complete a survey. Libraries appear to have ballpark estimates for how long it should take to complete their surveys. If the time it takes participants to complete the survey in the pilot tests exceeds this figure, questions might be omitted. The survey instructions include the estimated time required to complete the survey.

DLF respondents reported using different approaches to distribute or provide access to surveys, based on the sampling method and survey format. For example, when recruiting volunteers to take Web-based surveys, the survey might automatically pop up when users display the library home page or click the exit button on the online public access catalog (OPAC). Alternatively, a button or link on the home page might provide access to the survey. Posters or flyers might advertise the URL of a Web-based survey or, if a more carefully selected sample is needed, an e-mail address to contact to indicate interest in participating. Paper surveys may be made available in trays or handed to library users. With more carefully selected sample populations, e-mail containing log-in information to do a Web-based

survey, or the e-mail or paper survey itself, is sent to the targeted sample. Paper surveys can be distributed as e-mail enclosures or via campus or U.S. mail. DLF respondents indicated that all of these methods worked well.

Libraries use spreadsheet or statistical software to analyze the quantitative responses to surveys. Cross-tabulations are conducted to discover whether different user groups responded to the questions differently; for example, to discover whether the priorities of undergraduate students are different from those of graduate students or faculty. Some libraries compare the distribution of survey respondents with the demographics of the campus to determine whether the distribution of user groups in their sample is representative of the campus population. A few libraries have used content analysis software to analyze the responses to open-ended questions.

2.1.4. Who Uses Survey Results? How Are They Used?

Libraries share survey results with the people empowered to decide how those results will be applied. The formality of the survey and the sample size also determine who will see the results and participate in interpreting them and determining how they will be used. High-profile, potentially contentious survey topics or research purposes tend to be treated more formally. They entail the use of larger samples and generate more interest. Survey results of user satisfaction with the library Web site might be presented to the library governing council, which will decide how the results will be used. Data from more informal surveys might be shared strictly within the department that conducted the survey. For example, the results of a survey of user satisfaction with the laptop-lending program might be presented to the department, whose members will then decide whether additional software applications should be provided on the laptops. Striking or significant results from a survey of any size seem to bubble up to the attention of library administrators, particularly if follow-up might have financial or operational implications or require interdepartmental cooperation. For example, results of a survey of reference service that suggest that users would be better served by longer reference desk hours or staffing with systems office personnel in addition to reference librarians should be brought to the addition of library administration. Survey data might also be shared with university administrators, faculty senates, library advisory boards, and similar groups., to win or bolster support for changing directions in library strategic planning or to support requests for additional funding. Multiyear trends are often included in annual reports. The results are also presented at conferences and published.

Although survey results often confirm expectations and validate what the library is doing, sometimes the results are surprising. In this case, they may precipitate changes in library services, user interfaces, or plans. The results of the DLF survey indicate the following applications of survey data:

- Library administrators have used survey results to inform budget requests and secure funding from university administrators for

electronic resources and library facilities.

- Library administrators and middle managers have used survey results to guide reallocation of resources to better meet user needs and expectations. For example, low-priority services have been discontinued. More resources have been put into improving high-priority services with low satisfaction ratings or into enhancing existing services and tools or developing new ones.
- Collection developers have used survey results to inform investment decisions—for example, to decide which vendor's Modern Language Association (MLA) bibliography to license; whether to license a product after the six-month free trial period; or whether to drop journal titles, keep the titles in both print and electronic format, or add the journals in electronic format. Developers have also used survey data to inform collection-development decisions, for example, to set priorities for content to be digitized for inclusion in local collections or to decide whether to continue to create and collect analog slides rather than move entirely to digital images.
- Service providers, such as reference, circulation, and resource sharing (interlibrary loan [ILL] and document delivery) departments, have used survey results to identify problem areas and formulate steps to improve service quality in a variety of ways, for example, by reducing turnaround time for ILL requests, solving problems with network ports and dynamic host assignments for loaner laptops, helping users find new materials in the library, improving staff customer service skills, assisting faculty in the transition from traditional to electronic reserves, and developing or revising instruction in the use of digital collections, online finding aids, and vendor products.
- Developers have used survey results to set priorities and inform the customization or development of user interfaces for the OPAC, the library Web site, local digital collections, and online exhibits. Survey results have guided the revision of Web site vocabulary, the redesign of navigation and content of the library Web site, and the design of templates for personalized library Web pages. They have also been used to identify online exhibits that warrant upgrading.
- Survey results have been used to inform or establish orientation, technical competencies, and training programs for staff, to prepare reports for funding agencies, and to inform a Request for Proposals from ILS vendors.
- A multilibrary organization has conducted surveys to assess the need for original cataloging, the use of shared catalog records and vendor records, the standards for record acceptance (without local changes), and the applicability of subject classifications to library Web pages—all to inform plans for the future and ensure the appropriate allocation of cataloging resources.

DLF respondents mentioned that survey results often fueled discussion of alternative ways to solve problems identified in the survey. For example, when users report that they want around-the-clock

access to library facilities, libraries examine student wages (since students provide most of the staffing in libraries during late hours) and management of late-night service hours. When users complain that use of the library on a campus with many libraries is unnecessarily complicated, libraries explore ways to reorganize collections to reduce the number of service points. When users reveal that the content of e-resources is not what they expect, libraries evaluate their aggregator and document delivery services.

2.1.5. What Are the Issues, Problems, and Challenges With Surveys?

2.1.5.1. The Costs and Benefits of Different Types of Surveys

DLF respondents agreed that general surveys are not very helpful. Broad surveys of library collections and services do provide baseline data and, if the same questions are repeated in subsequent surveys, offer longitudinal data to track changing patterns of use. However, such surveys are time-consuming and expensive to prepare, conduct, and interpret. Getting people to complete them is difficult. The results are shallow and require follow-up research. Some libraries believe the costs of such surveys exceed the benefits and that important usage trends can be tracked more cost-effectively using transaction log analysis. (See section 3.)

Point-of-use surveys that focus on a specific subject, tool, or product work as well as, or better than, general surveys. They are quicker to prepare and conduct, easier to interpret, and more cost-effective than broad surveys. However, they must be repeated periodically to assess trends, and they, too, frequently require follow-up research.

User satisfaction surveys can reveal problem areas, but they do not provide enough information to solve the problems. Service quality surveys, based on the gap model (which measures the "gap" or difference between users' perceptions of excellent service and their perceptions of the service they received), are preferred because they provide enough information to plan service improvements. Unfortunately, service quality surveys are much more expensive to conduct than user satisfaction surveys.

2.1.5.2. The Frequency of Surveys

Surveys are so popular that DLF respondents expressed concern about their number and frequency. Over-surveying can decrease participation and make it more difficult to recruit participants. When the number of completed surveys is very small, the results are meaningless. Conducting surveys as a way to market library resources might exacerbate the problem.

2.1.5.3. Composing Survey Questions

The success of a survey depends on the quality and precision of the questions asked—their wording, presentation, and appropriateness to the research purpose. In the absence of in-house survey expertise, adequate training, or consultation with an expert, library surveys

often contain ambiguous or inaccurate questions. In the worst cases, the survey results are meaningless and the survey must be entirely revised and conducted again the following year. More likely, the problem applies to particular questions rather than to the entire survey. For example, one DLF respondent explained that a survey conducted to determine the vocabulary to be used on the library Web site did not work well because the categories of information that users were to label were difficult to describe, particularly the category of "full-text" electronic resources. Developing appropriate and precise questions is the key reason for pilot testing survey instruments.

Composing well-worded survey questions requires a sense of what respondents know and how they are likely to respond. DLF respondents reported the following examples. A survey conducted to assess interface design based on heuristic principles did not work well, probably because the respondents lacked the knowledge and skills necessary to apply heuristic principles to interface design (see section 2.4.1.1). Surveys that ask respondents to specify the priority of each service or collection in a list yield results where everything is simply ranked either "high" or "low," which is not particularly informative. Similarly, surveys that ask respondents how often they use a service or collection yield results of either "always use" or "never use." Where it is desirable to compare or contrast collections or services, it is important to require users to rank the *relative* priority of services or collections and to rank the *relative* frequency of use. Otherwise, interpreting the results will be difficult.

Asking open-ended questions and soliciting comments can also be problematic. Many respondents will not take the time to write answers or comments. If they do, the information they provide can offer significant insights into user perceptions, needs, and expectations. However, analyzing the information is difficult, and the responses can be incomplete, inconsistent, or illegible. One DLF respondent reported having hundreds of pages of written responses to a large survey. Another respondent explained that he and his staff "spent lots of time figuring out how to quantify written responses." A few DLF libraries have attempted to automate the process using content analysis software, but none of them was pleased with the results. Perhaps the problem is trying to extract quantitative results from qualitative data. The preferred approach appears to be to limit the number of open-ended questions and analyze them manually by developing conceptual categories based on the content of the comments. Ideally, the categories would be mutually exclusive and exhaustive (that is, all the data fit into one of them). After the comments are coded into the categories, the gist would be extracted and, if possible, associated with the quantitative results of the survey. For example, do the comments offer any explanations of preferences or problems revealed in the quantitative data? The point is to ask qualitative questions if and only if you have the resources to read and digest the results and if your aims in conducting the survey are at least partly subjective and indicative, as opposed to precise and predictive.

2.1.5.4. Lack of Analysis or Application

Theoretically, the process is clear: prepare the survey, conduct the survey, analyze and interpret the results, decide how to apply them, and implement the plan. In reality, the process frequently breaks down after the survey is conducted, regardless of how carefully it was prepared or how many hundreds of respondents completed it. Many DLF respondents reported surveys whose results were never analyzed. Others reported that survey results were analyzed and recommendations made, but nothing happened after that. No one knew, or felt comfortable enough to mention, who dropped the ball. No one claimed that changes in personnel were instrumental in the failure to analyze or apply the survey results. Instead, they focused on the impact this has on the morale of library staff and users. Conducting research creates expectations; people expect results. Faculty members in particular are not likely to participate in library research studies if they never see results. Library staff members are unlikely to want to serve on committees or task forces formed to conduct studies if the results are never applied.

The problem could be loss of momentum and commitment, but it could also be lack of skill. Just as preparing survey questions requires specific skills, so too do analysis, interpretation, and application of survey results. Libraries appear to be slow in acquiring the skills needed to use survey data. The problem is exacerbated when survey results conflict with other data. For example, a DLF respondent reported that their survey data indicate that users do not want or need reference service, even though the number of questions being asked at the reference desk is increasing. Morale takes a hit if no concrete next steps can be formulated from survey results or if the data do not match known trends or anecdotal evidence. In such cases, the smaller the sample, the more likely the results will be dismissed.

2.1.5.5. Lack of Resources or Comprehensive Plans

Paper surveys distributed to a statistically significant sample of a large university community can cost more than $10,000 to prepare, conduct, and analyze. Many libraries cannot afford or choose not to make such an investment. Alternative formats and smaller samples seem to be the preferred approach; however, even these take a considerable amount of time. Furthermore, surveys often fail to provide enough information to enable planners to solve the problems that have been identified. Libraries might not have the human and financial resources to allocate to follow-up research, or they could simply have run out of momentum. The problem could also be a matter of planning. If the research process is not viewed from conception through application of the results and follow-up testing, the process could likely halt at the point where existing plans end.

2.2. Focus Groups

2.2.1. What Is a Focus Group?

A focus group is an exploratory, guided interview or interactive conversation among seven to ten participants with common interests or characteristics.[3] The purpose of a focus group is to test hypotheses; reveal what beliefs the group holds about a particular product, service, or opportunity and why; or to uncover detailed information about complex issues or behaviors from the group's perspective. Focus group studies entail several such group conversations to identify trends and patterns in perception across groups. Careful analysis of the discussions reveals insights into how each group perceives the topic of discussion.

A focus group interview is typically one to two hours long. A trained moderator guides the conversation using five to ten predetermined questions or key issues prepared as an "interview guide." The questions are open-ended and noncommittal. They are simply stated and carefully articulated. The questions are asked in a specific sequence, but there are no predetermined response categories. The moderator clarifies anything that participants do not understand. The moderator may also ask probing follow-up questions to identify concepts important to the participants, pursue interesting leads, and develop and test hypotheses. In addition to the moderator, one or two observers take detailed notes.

Focus group discussions are audio- or videotaped. Audiotape is less obtrusive and therefore less likely to intimidate the participants. Participants who feel comfortable are likely to talk more than those who are not; for this reason, audiotape and well-trained observers are often preferred to videotape. The observers' notes should be so complete that they can substitute if the tape recorder does not work.

Focus groups are an effective and relatively easy way to gather insight into complex behavior and experience from the participants' perspective. Because they can reveal how groups of people think and feel about a particular topic and why they hold certain opinions, they are good for detecting changes in behavior. Participant responses can not only indicate what is new but also distinguish trends from fads. Interactive discussion among the participants creates synergy and facilitates recall and insight. A few focus groups can be conducted at relatively low cost. Focus group research can inform the planning and design of new programs or services, be it a means for evaluating existing programs or services, and facilitate the development of strategies for improvement and outreach. Focus groups are also helpful as prelude to survey or protocol research; they may be used to identify appropriate language, questions, or tasks, and as follow-up to survey or protocol research to get clarification or explanation of factors influencing survey responses or user behaviors. (Protocol research is discussed in section 2.3.)

[3] Much of the information in this section is taken from Chadwick, Bahr, and Albrecht 1984.

The quality of the responses to focus group questions depends on how clearly the questions are asked, the moderator's skills, and the participants' understanding of the goals of the study and what is expected of them. A skilled moderator is critical to the success of a focus group. Moderators must quickly develop rapport with the participant, remain impartial, and keep the discussion moving and focused on the research objectives. They should have background knowledge of the discussion topic and must be able to repress domineering individuals and bring everyone into the conversation. Before the focus group begins, the moderator should observe the participants and, if necessary, strategically seat extremely shy or domineering individuals. For example, outspoken, opinionated participants should be placed to the immediate left or right of the moderator and quiet-spoken persons must be placed at some distance from them. This enables the moderator to shut out the domineering person simply by turning his or her torso away from the individual. Moderators and observers must avoid making gestures (for example, head nodding) or comments that could bias the results of the study.

Moderators must be carefully selected, because attitude, gender, age, ethnicity, race, religion, and even clothing can trigger stereotypical perceptions in focus group participants and bias the results of the study. If participants do not trust the moderator, are uncomfortable with the other participants, or are not convinced that the study or their role is important, they can give incomplete, inaccurate, or biased information. To facilitate discussion, reduce the risk of discomfort and intimidation, and increase the likelihood that participants will give detailed, accurate responses to the focus group questions, focus groups should be organized so that participants and, in some cases, the moderator are demographically similar.

The selection of demographic participant groupings and focus group moderator should be based on the research purpose, the sensitivity of the topic, and an understanding of the target population. For example, topics related to sexual behavior or preferences suggest conducting separate focus groups for males and females in similar age groups with a moderator of the same age and gender. When the topic is not sensitive and the population is diverse, the research purpose is sufficient to determine the demographic groupings for selecting participants. For example, three focus groups—for undergraduate students, graduate students, and faculty—could be used to test hypotheses about needs or expectations for library resources among these groups. Mixing students and faculty could intimidate undergraduates. Although homogeneity is important, focus group participants should be sufficiently diverse to allow for contrasting opinions. Ideally, the participants do not know one another. This is because if they do, they tend to form small groups within the focus group and make it harder for the moderator to manage.

The primary disadvantage of focus groups is that participants may give false information to please the moderator, stray from the topic, be influenced by peer pressure, or seek a consensus rather than explore ideas. A dominating or opinionated participant can make

more reserved participants hesitant to talk, which could bias the results. In addition, data gathered in focus groups can be difficult to evaluate because such information can be chaotic, qualitative, or emotional rather than objective. The findings should be interpreted at the group level. The small number of participants and frequent use of convenience sampling severely limit the ability to generalize the results of focus groups, and the results cannot be generalized to groups with different demographic characteristics. However, the results are more intelligible and accessible to lay audiences and decision makers than are complex statistical analyses of survey data.

A final disadvantage of focus groups is that they rely heavily on the observational skills of the moderator and observer(s), who will not see or hear everything that happens, and will see or hear even less when they are tired or bored. How the moderators or observers interpret what they see and hear depends on their point of reference, cultural bias, experience, and expectations. Furthermore, observers adjust to conditions. They may eventually fail to recognize language or behaviors that become commonplace in a series of focus groups. In addition, human beings cannot observe something without changing it. The Heisenberg principle states that any attempt to get information out of a system changes it. In the context of human subjects research, this is called the Hawthorne or "guinea pig" effect. Being a research subject changes the subject's behavior. Having multiple observers can compensate for many of these limitations and increase the accuracy of observational studies, but it can also further influence the behaviors observed. The best strategy is to articulate the specific behaviors or aspects of behavior to be observed before conducting the study. Deciding, on the basis of the research objectives, what to observe and how to record the observations, coupled with training the observers, facilitates systematic data gathering, analysis of the research findings, and the successful completion of observational studies.

2.2.2. Why Do Libraries Conduct Focus Groups?

More than half of the DLF respondents reported conducting focus groups. They chose to conduct focus groups rather than small, targeted surveys because focus groups offer the opportunity to ask for clarification and to hear participants converse about library topics. Libraries have conducted focus groups to assess what users do or want to do and to obtain information on the use, effectiveness, and usefulness of particular library collections, services, and tools. They have also conducted focus groups to verify or clarify the results from survey or user protocol research, to discover potential solutions to problems identified in previous research, and to help decide what questions to ask in a survey. One participant reported conducting focus groups to determine how to address practical and immediate concerns in implementing a grant-funded project.

Data gathered from focus groups are used to inform decision making, strategic planning, and resource allocation. Focus groups have the added benefit of providing good quotations that are effec-

tive in public relations publications and presentations or proposals to librarians, faculty, university administrators, and funders. Several DLF respondents observed that a few well-articulated comments from users in conjunction with quantitative data from surveys or transaction log analysis can help make a persuasive case for changing library practice, receiving additional funding, or developing new services or tools.

2.2.3. How Do Libraries Conduct Focus Groups?

DLF respondents reported conducting focus groups periodically. Questions asked in focus groups, unlike those included in surveys, are not repeated; they are not expected to serve as a basis for assessing trends over time. The decision to convene a focus group appears to be influenced by the organization of the library and the significance or financial implications of the decision to be informed by the focus group data. For example, in a library with an established usability program or embedded culture of assessment (including a budget and in-house expertise), a unit head can initiate focus group research. If the library must decide whether to purchase an expensive product or undertake a major project that will require the efforts of personnel throughout the organization, a larger group of people might be involved in sanctioning and planning the research and in approving the expenditure to conduct it.

Once the decision has been made to conduct focus groups, one or more librarians or staff prepare the interview questions, identify the demographic groups appropriate for the research purpose, determine how many focus groups to conduct, decide how to recruit participants, and plan the budget and timetable for gathering, analyzing, interpreting and applying the data.

Focus group questions should be pilot tested with a group of users and revised on the basis of the test results to solve problems with vocabulary, wording, or the sequence of questions, and to ensure that the questions can be discussed in the allotted time. However, few DLF respondents reported testing focus group questions. More likely, the questions are simply reviewed by other librarians and staff before conducting the study. Questions are omitted or reorganized during the initial focus group session, on the basis of time constraints and the flow of the conversation. The revised list of questions is used in subsequent focus groups.

DLF libraries have used e-mail, posters, and flyers to recruit participants for focus group studies. The invitations to prospective participants briefly describe the goals and significance of the study, the participants' role in the study, what is expected of them, how long the groups will last, and any token of appreciation that will be given to the participants. Typically, focus groups are scheduled for 60 to 90 minutes. If food is provided during the focus group, a 90-minute session is preferred. When efforts fail to recruit at least six participants for a group, some libraries have conducted individual interviews with the people they did recruit.

In addition to preparing interview questions and recruiting and scheduling participants, focus group preparation entails the following:

- Recruiting, scheduling, and training a moderator and observer(s) for each focus group
- Scheduling six to twelve (preferably seven to ten) participants in designated demographic groups, and sending them a reminder a week or a few days before the focus group
- Scheduling an appropriate room for each focus group. DLF respondents offered the following cautions:
  - Make sure that the participants can easily find the room. Put up signs if necessary.
  - Beware of construction or renovation nearby, the sound of heating or air-conditioning equipment, and regularly scheduled noise makers (for example, a university marching band practice on the lawn outside).
  - Ensure that there are sufficient chairs in the room to comfortably seat the participants, moderator, and observer(s) around a conference table.
  - If handouts are to be distributed, for example, for participants to comment on different interface designs, be sure that the table is large enough to spread out the documents.
- Ordering food if applicable
- Photocopying the focus group questions for the moderator and observer(s)
- Testing the audio- or videotape equipment and purchasing tapes

The focus group moderator or an observer typically arrives at the room early, adjusts the light and temperature in the room, arranges the chairs, and retests and positions the recording equipment. If audiotape is used, a towel or tablet is placed under the recording device to absorb any table vibrations. When the participants arrive, the moderator thanks them for participating, introduces and explains the roles of moderator and observer, reiterates the purpose and significance of the research, confirms that their anonymity will be preserved in any discussion or publication of the study, and briefly describes the ground rules and how the focus group will be conducted. The introductory remarks emphasize that the goal of the study is not for the participants to reach consensus, but to express their opinions and share their experiences and concerns. Disagreement and discussion are invited. Sometimes the first question is asked round-robin, so that each participant responds and gets comfortable talking. Subsequent questions are answered less formally, more conversationally. The moderator asks the prepared questions and may ask undocumented, probing questions or invite further comments to better understand what the participants are saying and test relevant hypotheses that surface during the discussion. For example, "Would you explain that further?" or "Please give me an example." The moderator uses verbal and body language to invite comments from shy or quiet participants and to discourage domineering individuals from turning dialogue into monologue. If participants ask questions unre-

lated to the research purpose, the moderator indicates that the question is outside the scope of the topic under discussion, but that he or she will be happy to answer it after the focus group is completed. Observers have no speaking roles.

When the focus group is over, the moderator thanks the participants and might give them a token of appreciation for their participation. The moderator may also answer any questions the participants have about the study, the service or product that was the focus of the study, or the library in general. Observer notes and tapes are labeled immediately with the date and number of the session.

Libraries might or might not transcribe the focus group tapes. Some libraries believe the cost of transcribing exceeds the benefits of having a full transcription. One DLF respondent explained that clerical help is typically unfamiliar with the vocabulary or acronyms used by focus group participants and therefore cannot accurately transcribe the tapes. This means that a professional must also listen to the tapes and correct the transcriptions, which significantly increases the cost of the study. When the tapes are transcribed, a few libraries have used content analysis software to analyze the transcriptions, but they have not been pleased with the results, perhaps because the software attempts to conduct a quantitative analysis of qualitative data. Even when the tapes are not transcribed, at least one person listens to them carefully and annotates the notes taken by observers.

Analysis of focus group data is driven by the research purpose. Ideally, at least two people analyze the data—the moderator and observer—and there is high interrater reliability. With one exception, DLF respondents did not discuss the process of analyzing focus group data in detail. They talked primarily about their research purpose, what they learned, and how they applied the results. Participants who mentioned a specific method of data analysis named content analysis, but they neither described how they went about it nor specified who analyzed the data. No one offered an interrater reliability factor. Only one person provided details about the data analysis and interpretation. This person explained that the moderator analyzed the focus group data by using content analysis to cluster similar concepts, examining the context in which these concepts occurred, looking for changes in the focus group participants' position based on the discussion, weighting responses based on the specificity of the participants' experience, and looking for trends or ideas that cut across one or more focus group discussions. The overall impression from the DLF survey is that focus group data are somehow examined by question and user group to identify issues, problems, preferences, priorities, and concepts that surface in the data. The analyst prepares a written summary of significant findings from each focus group session, with illustrative examples or quotations from the raw data. The summaries are examined to discern significant differences among the groups or to determine whether the data support or do not support hypotheses being tested.

2.2.4. Who Uses Focus Group Results? How Are They Used?

Decisions as to who applies the results of focus group research and how it is applied depend on the purpose of the research, the significance of the findings, and the organization of the library. For example, the results of focus groups conducted to inform redesign of the library Web site were presented to the Web Redesign Committee. The results of focus groups conducted to assess the need for and use of electronic resources were presented to the Digital Library Initiatives Department. The larger the study, the more attention it seems to draw. Striking or significant results come to the attention of library administrators, especially if potential next steps have financial or operation implications or require interdepartmental cooperation. For example, if the focus group results indicate that customer service training is required or that facilities must be improved to increase user satisfaction, the administrator should be informed. Focus groups provide excellent quotations in support of cases being presented to university administrators, faculty senates, and deans' councils to gain support for changing library directions or receiving additional funding. The results are also presented at conferences and published in the library literature.

The results of the DLF study indicate that focus group data have been used to

- Clarify or explain factors influencing survey responses, for example, to discover reasons for undergraduate students' declining satisfaction with the library
- Determine questions to ask in survey questionnaires, tasks to be performed in protocols, and the vocabulary to use in these instruments
- Identify user problems and preferences related to collection format and system design and functionality
- Confirm hypotheses that user expectations and perceived needs for a library Web site differ across discipline and user status
- Confirm user needs for more and better library instruction
- Confirm that faculty are concerned that students cannot judge the quality of resources available on the Web and do not appreciate the role of librarians in selecting quality materials
- Target areas for fundraising
- Identify ways to address concerns in grant-funded projects

In addition, results from focus group research have been used to inform processes that resulted in

- Canceling journal subscriptions
- Providing needed information to faculty
- Redesigning the library Web site, OPAC, or other user interface
- Providing personalized Web pages for library users
- Sending librarians and staff to customer service training
- Eliminating a high-maintenance method of access to e-journals
- Planning the direction and development priorities for the digital library, including the scope, design, and functionality of digital library services

- Planning and allocating resources to market library collections and services continuously
- Creating a Distance Education Department to integrate distance learning with library services
- Renovating library facilities

2.2.5. What Are the Issues, Problems, and Challenges with Focus Groups?

2.2.5.1. Unskilled Moderators and Observers

If the moderator of a focus group is not well trained or has a vested interest in the research results, the discussion can easily go astray. Without proper facilitation, some individuals can dominate the conversation, while others may not get the opportunity to share their views. Faculty in particular can be problematic subjects. They frequently have their own agendas and will not directly answer the focus group questions. A skilled, objective moderator equipped with the rhetorical strategies and ability to keep the discussion on track, curtail domineering or rambling individuals, and bring in reticent participants is a basic requirement for a successful focus group.

Similarly, poor observer notes can hinder the success of a focus group. If observers do not know what comments or behaviors to observe and record, the data will be difficult, if not impossible, to analyze and interpret. The situation worsens if several observers attend different focus group sessions and record different kinds of things. Decisions should be made before conducting the focus groups to ensure that similar behaviors are observed and recorded during each focus group session. The following list can serve as a starting point for this discussion (Marczak and Sewell).

- Characteristics of the focus group participants
- Descriptive phrases or words used by participants in response to the key questions
- Themes in the responses to the key questions
- Subthemes held by participants with common characteristics
- Indications of participant enthusiasm or lack of enthusiasm
- Consistency or inconsistency between participant comments and observed behaviors
- Body language
- The mood of the discussion
- Suggestions for revising, eliminating, adding questions in the future

2.2.5.2. Interpreting and Using the Data

A shared system of categories for recording observations will simplify the analysis and interpretation of focus group data. No DLF respondent mentioned establishing such a system before conducting a focus group study. Imposing a system after the data have been gathered significantly complicates interpreting the findings. The difficulty of interpreting qualitative data from a focus group study can lead to disagreement about the interpretation and delay preparation of the results. The limited number of participants in a typical focus

group study, and the degree to which they are perceived to be representative of the target population, exacerbate the difficulty of interpreting and applying the results. The greater the time lapse between gathering the data and developing plans to use the data, the greater the risk of loss of momentum and abandonment of the study. The results of the DLF study suggest that the problem worsens if the results are presented to a large group within the library and if the recommended next steps are unpopular with or counterintuitive to librarians.

2.3. User Protocols

2.3.1. What Is a User Protocol?

A user protocol is a structured, exploratory observation of clearly defined aspects of the behavior of an individual performing one or more designated tasks. The purpose of the protocol is to gather in-depth insight into the behavior and experience of a person using a particular tool or product. User protocol studies include multiple research subjects to identify trends or patterns of behavior and experience. Data gathered from protocols provide insight into what different individuals do or want to do to perform specific tasks.

Protocol studies usually take 60 to 90 minutes per participant. The protocol is guided by a list of five to ten tasks (the "task script") that individuals are expected to perform. Each participant is asked to think aloud while performing the designated tasks. The task script is worded in a way that tells the user *what* tasks to accomplish (for example, "Find all the books in the library catalog published by author Walter J. Ong before 1970), but not told *how* to accomplish the tasks using the particular tool or product involved in the study. Discovering whether or how participants accomplish the task is a typical goal of protocol research. A facilitator encourages the participants to think aloud if they fall silent. The facilitator may clarify what task is to be performed, but not how to perform it.

The participant's think-aloud protocol is audio- or videotaped, and one or two observers take notes of his or her behavior. Some researchers prefer audiotape because it is less obtrusive. Experts in human-computer interaction (HCI) prefer videotape. In HCI studies, software can be used to capture participant keystrokes.

Protocols are very strict about the observational data to be collected. Before the study, the protocol author designates the specific user comments, actions, and other behaviors that observers are to record. The observers' notes should be so complete that they can substitute for the audiotape, should the system fail. In HCI studies, observer notes should capture the participant's body language, selections from software menus or Web pages, what the user apparently does or does not see or understand in the user interface, and, depending on the research goals, the speed and success (or failure) of task completion. Employing observers who understand heuristic principles of good design facilitates understanding the problems us-

ers encounter, and therefore the recording of what is observed and interpretation of the data.

User protocols are an effective method to identify usability problems in the design of a particular product or tool, and often the data provide sufficient information to enable the problems identified to be solved. These protocols are less useful to identify what works especially well in a design. Protocols can reveal the participant's mental model of a task or the tool that he or she is using to perform the task. Protocols enable the behavior to be recorded as it occurs and do not rely on the participants' memories of their behaviors, which can be faulty. Protocols provide accurate descriptions of situations and, unlike surveys, can be used to test causal hypotheses. Protocols also provide insights that can be tested with other research methods and supplementary data to qualify or help interpret data from other studies.

For protocols to be effective, participants must understand the goals of the study, appreciate their role in the study, and know what is expected of them. The selection of participants should be based on the research purpose and an understanding of the target population. Facilitators and observers must be impartial and refrain from providing assistance to struggling or frustrated participants. However, a limit can be set on how much time participants may spend trying to complete a task, and facilitators can encourage participants to move to the next task if the time limit is exceeded. Without a time limit, participants can become so frustrated trying to complete a task that they abandon the study. In HCI studies, it is essential that the participants understand it is the *software* that is being tested, not their skill in using it.

The primary disadvantage of user protocols is that they are expensive. Protocols require at least an hour per participant, and the results apply only to the particular product or tool being tested. In addition, protocol data can be difficult to evaluate, depending on whether the research focuses on gathering qualitative information (for example, the level of participant frustration) or quantitative metrics (for example, success rate and speed of completion). The small number of participants and frequent use of convenience sampling limit the ability to generalize the results of protocol studies to groups with different demographic characteristics or to other products or tools. Furthermore, protocols suffer from the built-in limitations of human sensory perception and language, which affect what the facilitator and observer(s) see and hear and how they interpret and record it.

2.3.2. Why Do Libraries Conduct User Protocols?

Half of the DLF respondents reported conducting or planning to conduct user protocols. With rare exception, libraries appear to view think-aloud protocols as the premier research method for assessing the usability of OPACs, Web pages, local digital collections, and vendor products. Protocol studies are often precipitated or informed by the results of previous research. For example, focus groups, surveys,

and heuristic evaluations can identify frequently performed or suspected problematic tasks to be included in protocol research. (Heuristic evaluations are discussed in section 2.4.1.1.)

Libraries participating in the DLF study have conducted think-aloud protocols to

- Identify problems in the design, functionality, navigation, and vocabulary of the library Web site or user interfaces to different products or digital collections
- Assess whether efforts to improve service quality were successful
- Determine what information to include in a Frequently Asked Questions (FAQ) database and the design of access points for the database

One DLF respondent reported plans to conduct a protocol study of remote storage robotics.

2.3.3. How Do Libraries Conduct User Protocols?

DLF respondents reported conducting user protocols when the results of previous research or substantial anecdotal evidence indicated that there were serious problems with a user interface or when a user interface was being developed as part of a grant-funded project, in which case the protocol study is described in the grant proposal. When protocols are conducted to identify problems in a user interface, often they are repeated later, to see whether the problems were solved in the meantime. In the absence of an established usability-testing program and budget, the decision to conduct protocols can involve a large group of people because of the time and expense of conducting such research.

After the decision has been made to conduct user protocols, one or more librarians or staff members prepare the task script, choose the sampling method, identify the demographic groups appropriate for the research purpose, determine how many participants to recruit in each group, decide how to recruit them, recruit and schedule the participants, and plan the budget and timetable for gathering, analyzing, interpreting and applying the data. Jakob Nielsen's research has shown that four to six subjects per demographic group is sufficient to capture most of the information that could be discovered by involving more subjects. Beyond this number, the cost exceeds the benefits of conducting more protocols (Nielsen 2000). Sometimes protocols are conducted with only two or three subjects per user group because of the difficulty of recruiting research subjects.

DLF libraries immediately follow user protocol sessions with a brief survey or interview to gather additional information from each participant. This information helps clarify the user's behavior and provides some sense of the user's perception of the severity of the problems encountered with the user interface. One or more people prepare the survey or interview questions. In addition, some libraries prepare a recording sheet that observers use to structure their observations and simplify data analysis. Some also prepare a written facilitator guide that outlines the entire session.

DLF libraries pilot test the research instruments with at least one user and revise them on the basis of the test results. Pilot testing can help solve problems with the vocabulary, wording, or sequencing of protocol tasks or survey questions; it also can target ways to refine the recording sheet to facilitate rapid recording of observations. Pilot testing also enables the researcher to ensure that the protocol and follow-up research can be completed in the time allotted.

DLF libraries have used e-mail, posters, and flyers to recruit participants for user protocol studies. The recruitment information briefly describes the goals and significance of the research, the participants' role, and what is expected of them, including the time it will take to participate and any token of appreciation that will be given to the participants. Other than preparing the instruments and recruiting participants, preparation for a user protocol study closely resembles preparation for a focus group. It involves the following steps:

- Recruiting, scheduling, and training a facilitator and one or more observers; in some cases, the facilitator is the sole observer
- Scheduling the participants and sending them a reminder a week or a few days before the protocol
- Scheduling a quiet room; protocol studies have been conducted in offices, laboratories, or library settings.
- If necessary, ordering computer or videotape equipment to be delivered a half hour before the protocol is to begin
- Photocopying the research instruments
- Testing the audio- or videotape equipment and purchasing tapes

The facilitator or an observer arrives at the room early, adjusts the light and temperature in the room, arranges the chairs so that the facilitator and observers can see the user's face and the computer screen, and tests and positions the recording equipment. If audiotape is used, a towel or tablet is placed under the recording device to absorb any table vibrations. The audiotape recorder is positioned close enough to the user to pick up his or her comments, but far enough away from the keyboard to avoid capturing each key click. If computer or videotape equipment must be delivered to the room, someone must arrive at the room extra early to confirm delivery, be prepared to call if it is not delivered, test the computer equipment, and allow time for replacement or software reinstallation if something is not working.

Though HCI experts recommend videotape, all but one of the DLF libraries reported using audiotape to record user protocols. The library that used videotape observed that the camera made users uncomfortable and the computer screen did not record well, so the group used audiotape instead for the follow-up protocols. Few DLF libraries have the resources or facilities to videotape their research, and the added expense of acquiring these might also be a deterrent to using videotape.

When participants arrive, the facilitator thanks then for participating, explains the roles of facilitator and observer(s), reiterates the purpose and significance of the research, confirms that anonymity

will be preserved in any discussion or publication of the study, and describes the ground rules and how the protocol will be conducted. The facilitator emphasizes that the goal of the study is to test the software, not the user. The facilitator usually reminds participants multiple times to think aloud. For example, "What are you thinking now?" or "Please share your thoughts." Observers have no speaking role.

DLF libraries immediately followed protocol sessions with brief interviews or a short survey to capture additional information and give participants the opportunity to clarify what they did in the protocol, describe their experience, and articulate expectations they had about the task or the user interface that were not met. Protocol research is sometimes followed up with focus groups or surveys to confirm the findings with a larger sample of the target population.

When the protocol is over, the facilitator thanks the participant and usually gives him or her a token of appreciation. The facilitator also answers any questions the participant has. Observer notes and tapes are labeled immediately.

DLF libraries might or might not transcribe protocol tapes for the same reasons they do or do not transcribe focus group tapes. If the tapes are not transcribed, at least one person listens to them and annotates the observer notes. With two exceptions, DLF respondents did not discuss the process of analyzing, interpreting, and figuring out how to apply the protocol results, although several did mention using quantitative metrics. They simply talked about significant applications of the results. The two cases that outlined procedures for analyzing, interpreting, and applying results merit examination:

- Case one: The group responsible for conducting the protocol study created a table of observations (based on the protocol data), interpretations, and accompanying recommendations for interface redesign. The recommendations were based on the protocol data and the application of Jakob Nielsen's 10 heuristic principles of good user interface design (Nielsen, no date). The group assessed how easy or difficult it would be to implement each recommendation and plotted a continuum of recommendations based on the difficulty, cost, and benefit of implementing them. The cost-effective recommendations were implemented.
- Case two: When protocol data identified many problems and yielded a high failure rate for task completion, the group responsible for the study did the following:
  - Determined the severity of each problem on the basis of its frequency and distribution across users, whether it prevented users from successfully completing a task, and the user's assessment of the severity of the problem, which was gathered in a follow-up survey.
  - Formulated alternative potential solutions to the most severe problems on the basis of the protocol or follow-up survey data and heuristic principles of good design.
  - Winnowed the list of possible solutions by consulting programmers and doing a quick-and-dirty cost-benefit analysis. Prob-

lems that can be fixed at the interface level are often less expensive to fix than those that require changes in the infrastructure.
- Recommended implementing the solutions believed to have the greatest benefit to users for the least amount of effort and expense.

The procedures in the two cases are similar, and although the other DLF respondents did not describe the process they followed, it could be that their processes resemble these. At least one other respondent reported ranking the severity of problems identified by protocol analysis to determine which problems to try to solve.

2.3.4. Who Uses Protocol Results? How Are They Used?

The results of the study suggest that who applies the results from user protocols and how the results are applied depend on the purpose of the research, the significance of the findings, and the organization of the library. The larger the study and the more striking its implications for financial and human resources, the more attention it draws in the library. Although the results of protocol studies are not always presented to university administrators, faculty senates, deans' councils, and similar groups; they might be presented at conferences and published in the library literature.

DLF libraries have used significant findings from protocol analysis to inform processes that resulted in the following:

- Customizing the OPAC interface, or redesigning the library Web site or user interfaces to local digital collections. Examples of steps taken based on protocol results include
  - rearranging a hierarchy
  - changing the order and presentation of search results
  - changing the vocabulary, placement of links, or page layout
  - providing more online help, on-screen instructions, or suggestions when searches fail
  - changing the labeling of images
  - changing how to select a database or start a new search
  - improving navigation
  - enhancing functionality
- Revising the metadata classification scheme for image or text collections
- Developing or revising instruction for how to find resources on the library Web site and how to use full-text e-resources and archival finding aids

The results of protocol studies have also been used to suggest revisions or enhancements to vendor products, to verify improvements in interface design and functionality, and to counter anecdotal evidence or suggestions that an interface should be changed.

2.3.5. What Are the Issues, Problems, and Challenges With User Protocols?

2.3.5.1. Librarian Assumptions and Preferences

Several DLF respondents commented that librarians can find it difficult to observe user protocols because they often have assumptions about user behavior or preferences for interface design that are challenged by what they witness. Watching struggling or frustrated participants and refraining from providing assistance run counter to the librarians' service orientation. Participants often ask questions during the protocol about the software, the user interface, or how to use it. Facilitators and observers must resist providing answers during the protocol. Librarians who are unable to do this circumvent the purpose of the research.

Librarians can also be a problem when it comes to interpreting and applying the results of user protocols. Those trained in social science research methods often do not understand or appreciate the difference between HCI user protocols and more rigorous statistical research. They may dismiss results that challenge their own way of thinking because they believe the research method is not scientific enough or the pool of participants is too small.

2.3.5.2. Lack of Resources and Commitment

User protocols require skilled facilitators, observers, and analysts and the commitment of human and financial resources. Requisite skills might be lacking to analyze, interpret, and persuasively present the findings. Even if the skills are available, there could be a breakdown in the processes of collecting, analyzing, and interpreting the data, planning how to use the findings, and implementing the plans, which could include conducting follow-up research to gather more information. Often the process is followed to the last stage, implementation, where Web masters, programmers, systems specialists, or other personnel are needed. These people can have other priorities. Human and financial resources or momentum can be depleted before all the serious problems identified have been solved. Limited resources frequently restrict implementation to only the problems that are cheap and easy to fix, which are typically those that appear on the surface of the user interface. Problems that must be addressed in the underlying architecture often are not addressed.

2.3.5.3. Interpreting and Using the Data

Effective, efficient analysis of data gathered in user protocols depends on making key decisions ahead of time about what behaviors to observe and how to record them. For example, if quantitative usability metrics are to be used, they must be carefully defined. If the success rate is to be calculated, what constitutes success? Is it more than simply completing a task within a set time limit? What constitutes partial success, and how is it to be calculated? Similar questions should be posed and answers devised for qualitative data gathering during the protocols. Otherwise, observer notes are chaotic and data analysis may be as difficult as is analyzing the responses to open-

ended questions in a survey. The situation worsens if different observers attend different protocols and record different kinds of things. Such key decisions should be made prior to conducting the study. If made afterward, they can result in significant lag time between data gathering and the presentation of plans to apply the results of the data analysis. The greater the lag time, the greater the risk of loss of momentum, which can jeopardize the entire effort.

2.3.5.4. Recruiting Participants Who Can Think Aloud

General problems and strategies for recruiting research subjects are discussed in section 4.2.1. DLF respondents reported difficulty in getting participants to think aloud. At least one librarian is considering conducting screening tests to ensure that protocol participants can think aloud. Enhancing the skills of the facilitator (through training or experience) and including a pretest task or two for the participants to get comfortable thinking aloud would be preferable to risking biasing the results of the study by recruiting only participants who are naturally comfortable thinking aloud.

2.4. Other Effective Research Methods

2.4.1. Discount Usability Research Methods

Discount usability research can be conducted to supplement more expensive usability studies. This informal research can be done at any point in the development cycle, but is most beneficial in the early stages of designing a user interface or Web site. When done at this time, the results of discount usability research can solve many problems and increase the efficiency of more formal testing by targeting specific issues and reducing the volume of data gathered. Discount usability research methods are not replacements for formal testing with users, but they are fruitful, inexpensive ways to improve interface design. In spite of these merits, few DLF libraries reported using discount methods. Are leading digital libraries not using these research methods because they are unaware of them or because they do not have the skills to use them?

2.4.1.1. Heuristic Evaluations

Heuristic evaluation is a critical inspection of a user interface conducted by applying a set of design principles as part of an iterative design process.[4] The principles are not a checklist, but conceptual categories or rules that describe common properties of a usable interface and guide close scrutiny of an interface to identify where it does not comply with the rules. Several DLF respondents referred to Nielsen's heuristic principles of good design, mentioning the following:

- Visibility of system status
- Match between system and real world

[4] See, for example, Nielsen 1994. Other chapters in the book describe other usability inspection methods, including cognitive walk-throughs.

- User control and freedom
- Consistency and standards
- Recognition rather than recall
- Flexibility and efficiency of use
- Aesthetics and minimalist design
- Error prevention
- Assistance with recognizing, diagnosing, and recovering from errors
- Help and documentation[5]

Heuristic evaluations can be conducted before or after formal usability studies involving users. They can be conducted with functioning interfaces or with paper prototypes (see section 2.4.1.2.). Applying heuristic principles to a user interface requires skilled evaluators. Nielsen recommends using three to five evaluators, including someone with design expertise and someone with expertise in the domain of the system being evaluated. According to his research, a single evaluator can identify 35 percent of the design problems in the user interface. Five evaluators can find 75 percent of the problems. Using more than five evaluators can find more problems, but at this point the cost exceeds the benefits (Nielsen 1994).

Heuristic evaluations take one to two hours per evaluator. The evaluators should work independently but share their results. An evaluator can record his or her own observations, or an observer may record the observations made by the evaluator. Evaluators follow a list of tasks that, unlike the task script in a user protocol, may indicate how to perform the tasks. The outcome from a heuristic evaluation is a compiled list of each evaluator's observations of instances where the user interface does not comply with good design principles. To guide formulating solutions to the problems, each problem identified is accompanied by a list of the design principles that are violated in this area of the user interface.

Heuristic evaluations have several advantages over other methods for studying user interfaces. No participants need to be recruited. The method is inexpensive, and applying even a few principles can yield significant results. The results can be used to expand or clarify the list of principles. Furthermore, heuristic evaluations are more comprehensive than think-aloud protocols are, because they can examine the entire interface and because even the most talkative participant will not comment on every facet of the interface. The disadvantages of heuristic evaluations are that they require familiarity with good design principles and interpretation by an evaluator, do not provide solutions to the problems they identify, and do not identify mismatches between the user interface and user expectations. Interface developers sometimes reject the results of heuristic evaluations because no users were involved.

[5] A brief description of these principles is available in Nielsen, no date.

A few DLF libraries have conducted their own heuristic evaluations or have made arrangements with commercial firms or graduate students to do them. The evaluations were conducted to assess the user-friendliness of commercially licensed products, the library Web site, and a library OPAC. In the process, libraries have analyzed such details as the number of keystrokes and mouse movements required to accomplish tasks and the size of buttons and links that users must click. The results of these evaluations were referred to as a "wake-up call" to improve customer service. It is unclear from the survey whether multiple evaluators were used in these studies or the study was conducted in-house, and whether the libraries have the interface design expertise to apply heuristic principles or conduct a heuristic evaluation effectively. Nevertheless, several DLF libraries reported using heuristic principles to guide redesign of a user interface.

2.4.1.2. Paper Prototypes and Scenarios

Paper prototype and scenario research resembles think-aloud protocols, but instead of having users perform tasks with a functioning system, this method employs sketches, screen prints, or plain text and asks users how they would use a prototype interface to perform different tasks or how they would interpret the vocabulary. For example, where would they click to find a feature or information? What does a link label mean? Where should links be placed? Paper prototypes and scenarios can also be a basis for heuristic evaluations.

Paper prototype and scenario research is portable, inexpensive, and easy to assemble, provided that the interface is not too complicated. Paper prototypes do not intimidate users. If it is used early in the development cycle, the problems identified can be rectified easily because the system has not been fully implemented. Paper prototypes are more effective than surveys to identify usability, navigation, functionality, and vocabulary problems. The disadvantage is that participants interact with paper interfaces differently than they do with on-screen interfaces; that is, paper gets closer scrutiny.

A few DLF respondents reported using paper prototype research. They have used it successfully to evaluate link and button labels and to inform the design of Web sites, digital collection interfaces, and classification (metadata) schemes. One library used scenarios of *horizontal* paper prototypes, which provide a conceptual map of the entire surface layer of a user interface, and scenarios of *vertical* paper prototypes, which cover the full scope of a feature, such as searching or browsing. This site experimented with using Post-it™ notes to display menu selections in a paper prototype study, and accordion-folded papers to imitate pages that would require scrolling. The notes were effective, but the accordion folds were awkward.

2.4.2. Card-Sorting Tests

Vocabulary problems can arise in any user study, and they are often rampant in library Web sites. A few respondents reported conducting research specifically designed to target or solve vocabulary prob-

lems, including card-sorting studies to determine link labels and appropriate groupings of links on their Web sites. Card-sorting studies entail asking individual users to

- Organize note cards containing service or collection descriptions into stacks of related information
- Label the stacks of related information
- Label the service and collection descriptions in each stack

Reverse card-sorting exercises have been used to test the labels. These exercises ask users what category (label) they would use to find which service or collection. Alternatively, the researcher can simply ask users what they would expect to find in each category, then show them what is in each category and ask them what they would call the category.

The primary problem encountered in conducting card-sorting tests is describing the collections and services to be labeled and grouped. Describing "full-text" e-resources appears to be particularly difficult in card-sorting exercises, and the results of surveys, focus groups, and user protocols indicate that users often do not understand what "full-text" means. Unfortunately, this is the term found on many library Web sites.

3. USAGE STUDIES OF ELECTRONIC RESOURCES

3.1. What Is Transaction Log Analysis?

Transaction log analysis (TLA) was developed about 25 years ago to evaluate system performance. Over the course of a decade, it evolved as a method to study unobtrusively interactions between online information systems and the people who use them. Today, it is also used to study use of Web sites. Researchers who conduct TLA rely on transaction monitoring software, whereby the system or Web server automatically records designated interactions for later analysis. Transaction monitoring records the type, if not the content, of selected user actions and system responses. For example, a user submits a query in the OPAC. Both the fact that a query was submitted and the content of that query could be recorded. In response, the system conducts a search and returns a list of results. Both the fact that results were returned and the number of results could be recorded. Transaction monitoring software often captures the date and time of these transactions, and the Internet Protocol (IP) address of the user. The information recorded is stored in an electronic file called a "transaction log." The contents of transaction logs are usually formatted in fields to facilitate quantitative analysis. Researchers analyze transaction logs to understand how people use online information systems or Web sites with the intention of improving their design and functionality to meet user needs and expectations. The analysis can be conducted manually or automatically, using software or a script to mine data in the logs and generate a report.

TLA is an effective method to study such activities as the frequency and sequence of feature use; system response times; hit rates; error rates; user actions to recover from errors; the number of simultaneous users; and session lengths. In the library world, if queries are logged, it can reveal why searches fail to retrieve results and suggest areas for collection development. If the IP addresses of users are logged, it can reveal whether the user is inside or outside of the library. The information extracted from transaction logs can be used to assess patterns of use and trends over time, predict and prepare for times of peak demand, project future system needs and capacities, and develop services or interfaces that support user actions. For TLA to be effective, transaction monitoring software must record meaningful transactions, and data mining must be driven by carefully articulated definitions and purposes.

TLA is an unobtrusive way to study user behavior, an efficient way to gather longitudinal usage data, and an effective way to detect discrepancies between what users say they do (for example in a focus group study) and what they actually do when they use an online system or Web site. Transaction log analysis is also a good way to test hypotheses; for example, to determine whether the placement or configuration of public computers (for example, at stand-up or sit-down stations) in the library affects user behavior.

The primary disadvantages of TLA are that extracting data can be time-consuming and the data can be difficult to interpret. Though systems and servers have been logging transactions for decades, they still do not incorporate software to analyze the logs. If analysis is to be conducted routinely over time, programmers must develop software or scripts to mine the data in transaction logs. If additional information is to be mined, someone must do it manually or the programmer must add this capability to the routine. Often, extracting the data requires discussion and definitions. For example, in stateless, unauthenticated systems such as the Web environment, what constitutes a user session with a Web-based collection or a virtual visit to the library Web site?

Even after the data have been mined, interpreting the patterns or trends discovered in the logs can be problematic. For example, are a large number of queries necessarily better than a small number of queries? What if users are getting better at searching and able to retrieve in a single query what it might have taken them several queries to find a few years ago? Are all searches that retrieve zero results failed searches? What if it was a known-item search and the user just wanted to know whether the library has the book? What constitutes a failed search? Zero results? Too many results? How many is too many? Meaning is contextual, but with TLA, there is no way to connect data in transaction logs with the users' needs, thoughts, goals, or emotions at the time of the transaction. Interpreting the data requires not only careful definitions of what is being measured but additional research to provide contextual information about the users.

A further disadvantage is that transaction logs can quickly grow to an enormous size. The data must be routinely moved from the

server where they are captured to the server where they are analyzed. Keeping log files over time, in case a decision is made to mine additional data from the files, results in massive storage requirements or offline storage that can impede data mining.

3.2. Why Do Libraries Conduct Transaction Log Analysis?

Most of the DLF respondents reported conducting TLA or using TLA data provided by vendors to study use of the library Web site, the OPAC and integrated library system (ILS), licensed electronic resources, and, in some cases, local digital collections and the proxy server. They have used TLA data from local servers to

- Identify user communities
- Identify patterns of use
- Project future needs for services and collections
- Assess user satisfaction
- Inform digital collection development decisions
- Inform the redesign and development of the library Web site
- Assess whether redesign of the library Web site or digital collection has had any impact on use
- Assess whether providing additional content on the library Web site or digital collection has any impact on use
- Target marketing or instruction efforts
- Assess whether marketing or instruction has any impact on use
- Drive examinations of Web page maintenance requirements
- Inform capacity planning and decisions about platform
- Plan system maintenance
- Allocate human and financial resources

Vendor-supplied TLA data from licensed electronic resources have been used to

- Help secure funding for additional e-resources from university administrators
- Inform decisions about what subscriptions or licenses to renew or cancel
- Inform decisions about which interface(s) to keep
- Determine how many ports or simultaneous users to license
- Assess whether instruction has any impact on use of an e-resource
- Determine cost per-use of licensed e-resources

3.3. How Do Libraries Conduct Transaction Log Analysis?

3.3.1. Web Sites and Local Digital Collections

Practices vary significantly across institutions—from no analysis to extensive analysis. DLF libraries track use of their Web sites and Web-accessible digital collections using a variety of homegrown, shareware, or commercial software. The server software determines what information is logged and therefore what data are available for

mining. The logging occurs automatically, but decisions concerning what data are extracted appear to be guided by library managers, administrators, or committees. As different questions are asked, different data are extracted to answer them. For example, as libraries adopt new measures for digital library use, Web server logs are being mined for data on virtual visits to the library. In some libraries a great deal of discussion is involved in defining such things as a "virtual visit." In other libraries, programmers are instructed to make their best guesstimate, explain what it is and why they chose it, and use it consistently in mining the logs. As with user studies, the more people involved in making these decisions, the longer it can take. The longer it takes, the longer the library operates without answers to its questions.

Many libraries do not use Web usage data because they do not know how to apply them or do not have the resources to apply them. Some libraries, however, are making creative use of transaction logs from the library Web site and local digital collections to identify user communities, determine patterns of use, inform decisions, assess user satisfaction, and measure the impact of marketing, instruction, interface redesign, and collection development. They do this by mining, interpreting, and applying the following data over time:

- Number of page hits
- Number and type of files downloaded
- Referral URLs (that is, how users get to a Web page)
- Web browser used
- Query logs from "Search this site" features
- Query logs from digital collection (image) databases
- Date and time of the transactions
- IP address or Internet domain of the user
- User IDs (in cases where authentication is required)

In addition, several libraries have begun to count "click throughs" from the library Web site to remote e-resources using a "count use mechanism." This mechanism captures and records user clicks on links to remote online resources by retrieving and logging retrieval of an intermediate Web page. The intermediate page is retrieved and replaced with the remote resource page so quickly that users do not notice the intermediate page. Writing a script to capture click throughs from the library Web site to remote resources is apparently simple, but the mechanism requires that the links (URLs) to all remote resources on the library Web site be changed to the URL of the intermediate page, which contains the actual URL of the remote resource. Libraries considering implementing a count use mechanism must weigh the cost of these massive revisions against the benefits.

The count use mechanism provides a consistent, comparable count of access to remote e-resources from the library Web site, and it is the only way to track use of licensed resources for which the vendor provides no usage statistics. The data, however, provide an incomplete and inaccurate picture of use of remote resources because users can bookmark resources rather than click through the library

Web site to get to them, and because the mechanism counts all attempts to get to remote resources, some of which fail because the server is down or the user does not have access privileges to the resource.

3.3.2. OPAC and Integrated Library Systems

Both OPAC and ILS log transactions, but different systems log different information; therefore, each enables analysis of different user activities. For example, some systems simply count different types of transactions. Others log additional information, such as the text of queries, the date and time of the transaction, the IP address and interface of the client machine, and a session ID, which can be used to reconstruct entire user sessions. Systems can provide an on-off feature to allow periodic monitoring and reduce the size of log files, which can grow at a staggering rate if many transactions and details are captured.

Integrated library systems provide a straightforward way for libraries to generate summary reports of such things as the number of catalog searches, the number of items circulated, the number of items used in-house, and the number of new catalog records added within a given period. Use of different interfaces, request features (for example, renewals, holds, recalls, or requests for purchases) and the ability to view borrowing records might also be tracked. This information is extracted from system transaction logs using routine reporting mechanisms provided by the vendor, or special custom report scripts developed either in-house or prepared as work for hire by the vendor for a fee. Customized reports are produced for funding agencies or in response to requests for data relevant to specific problems or pages (for example, subject pages or pathfinders). Often Web and ILS usage data are exported to other tools for further analysis, manipulation, or use; for example, circulation data and the number of queries are exported to spreadsheet software to generate trend lines. In rare cases, Web forms and functionality are provided for staff to generate ad hoc reports.

3.4. Who Uses the Results of Transaction Log Analysis? How Are They Used?

3.4.1. Web Sites and Local Digital Collections

Staff members generate monthly usage reports and distribute or make them available to all staff or to the custodians of the Web pages or digital collection. Overall Web site usage (page hits) or the 10 most heavily used pages might be included in a library's annual report. However, though usage reports are routinely generated, often the data languish without being used.

At institutions where the data are used, many different people use the data for many different purposes. Interface designers, system managers, collection developers, subject specialists, library administrators, and department heads all reap meaning and devise next steps from examining and interpreting the data. Page hits and refer-

ral URLs are used to construct usage patterns over time, understand user needs, and inform interface redesign. For example, frequently used Web pages are placed one to two clicks from the home page; infrequently used links on the home page are moved one to two clicks down in the Web site. Data on heavily used Web pages prompt consideration of whether to expand the information on these pages. Similarly, data on heavily used digital collections prompt consideration of expanding the collection. Subject specialists use the data to understand how people use their subject pages and pathfinders and revise their pages based on this understanding. Page hit counts also drive examination of page maintenance requirements with the understanding that low-use pages and collections should be low maintenance; high-use pages should be well maintained, complete, and up to date. Such assessments facilitate appropriate allocation of resources. Data on low-use or no-use pages can be used to target publicity campaigns. Cross-correlations of marketing efforts and usage statistics are performed to determine whether marketing had any measurable effects on use. Similarly, correlating interface redesign or expansion of content with usage statistics can determine whether redesign or additional content had any effect on use. Data on use of "new" items on the Web site are used to determine whether designating a resource as "new" had any measurable effects on use. Tracking usage patterns over time enables high-level assessments of user satisfaction. For example, are targeted user communities increasingly using the library Web site or digital collection? Do referral URLs indicate that more Web sites are linking to the library Web site or collection?

Query logs are also mined, interpreted, and applied. Frequent queries in "Search this site" logs identify resources to be moved higher in the Web site. Unsuccessful queries target needed changes in Web site vocabulary or content. Query logs from image databases are used to adjust the metadata and vocabulary of digital collections to match the vocabulary and level of specificity of users and to help decide whether the content and organization of digital collections are appropriate to user needs.

TLA also informs system maintenance and strategic planning. Time and date stamps enable the monitoring of usage patterns in the context of the academic year. Libraries have analyzed low-use times of day and day of week to determine good times to take Web servers down for maintenance. Page hits and data on the number and type of files downloaded month-to-month are used to plan load and capacity, to characterize consumption of system resources, to prepare for peak periods of demand, and to make decisions about platform and the appropriate allocation of resources.

Although the use of dynamic IP addresses makes identification of user communities impossible, libraries use static IP addresses and Internet domain information (for example, .edu, .com, .org, .net) in transaction logs to identify broad user communities. Libraries are defining and observing the behavior of different communities. Some libraries track communities of users inside or outside the library.

Some track on-campus, off-campus, or international user communities; others track communities in campus dormitories, libraries, offices, computer clusters, or outside the university. In rare cases, static IP addresses and locations are used to affiliate users with a particular school, department, or research center—recognizing that certain IP address locations, such as libraries, dormitories, and public computing clusters, reveal no academic affiliation of the users. Where users are required to authenticate (for example, at the proxy server), the authentication data are mapped to the library patron database to identify communities by school and user status (such as humanities undergraduate). If school and user status are known, some libraries conduct factor analysis to identify clusters of use by user communities.

Having identified user communities in the transaction logs, libraries then track patterns of use by different communities and the distribution of use across communities. For example, IP addresses and time and date stamps of click-through transactions are used to identify user communities and their patterns of using the library Web site to access remote e-resources. IP addresses and time and date stamps of Web site usage are used to track patterns of use inside and outside the libraries. The patterns are then used to project future needs for services and collections. For example, what percentage of use is outside the library? Is remote use increasing over time or across user groups? What percentage of remote use occurs in dormitories (undergraduate students)? What services and collections are necessary to meet the needs of remote users? Patterns of use per user community and resource are used to target publicity about digital collections or Web pages.

3.4.2. OPAC and Integrated Library Systems

OPAC and ILS usage data are used primarily to track trends and provide data for national surveys, for example, circulation per year or items cataloged per year. At some institutions, these data are used to inform decisions. OPAC usage statistics are used to determine usage patterns, customize the OPAC interface, and allocate resources. Seldom-used indexes are removed from the simple search screen and buried lower in the OPAC interface hierarchy. More resources are put into developing the Web interface than the character-based (telnet) interface because usage data show that the former is more heavily used. Libraries shopping for a new ILS frequently use the data to determine the relative importance of different features and required functionality for the new system.

In addition to mining data in transaction logs, some libraries extract other information from the ILS and export it to other tools. For example, e-journal data are exported from the ILS to a Digital Asset Management System (DAMS) to generate Web page listings of e-journals. The journal call numbers are used to map the e-journals to subject areas, and the Web pages are generated using Perl scripts and persistent URLs that resolve to the URLs of the remote e-journal sites. One site participating in the DLF survey routinely exports in-

formation from the ILS to a homegrown desktop reporting tool that enables staff to generate ad hoc reports.

3.4.3. Remote Electronic Resources

Library administrators use vendor-provided data on searches, sessions, or full-text use of remote e-resources to lobby for additional funding from university administrators. Data on selected, high-use e-resources might be included in annual reports. Collection developers use the data to determine cost per use of various products and to inform decisions about what subscriptions, licenses, or interfaces to keep or drop. Turn-away data are used to determine how many ports or simultaneous users to license, which could account for why so few vendors provide this information. Reference librarians use the data to determine whether product instruction has any impact on product use. Plans to promote particular products or to conduct research are developed on the basis of data identifying low-use products. Usage data indicate whether promoting a product has any impact on product use. Libraries that require authentication to use licensed resources, capture the authentication data, and map it to the patron database have conducted factor analysis to cluster the use of different products by different user communities. Libraries that compile all of their e-resource usage statistics have correlated digital input and output data to determine, for example, that 22 percent of the total number of licensed e-resources accounts for 70 percent of the total e-resource use.

3.5. What Are the Issues, Problems, and Challenges with Transaction Log Analysis?

3.5.1. Getting the Right (Comparable) Data and Definitions

3.5.1.1. Web Sites and Local Digital Collections

DLF respondents expressed concern that the most readily available usage statistics might not be the most valuable ones. Page hit rates, for example, might be relevant on the open Web, where sites want to document traffic for their advertisers, but on the library Web site, what do high or low hit rates really mean? Because Web site usage changes so much over time, comparing current and past usage statistics presents another challenge.

Despite the level of creative analysis and application of Web usage data at some institutions, even these libraries are not happy with the software they use to analyze Web logs. The logs are huge and analysis is cumbersome, sometimes exceeding the capacity of the software. Libraries are simultaneously looking for alternative software and trying to figure out what data are useful to track, how to gather and analyze the data efficiently, and how to present the data appropriately to inform decisions. Ideally, to facilitate comparisons, libraries want the same data on Web page use, the use of local databases or digital collections, and the use of commercially licensed databases and collections.

Libraries also want digital library usage statistics to be comparable with traditional usage statistics. For example, they want to count virtual visits to the library and combine this information with gate counts to get a complete picture of library use. Tracking virtual visits is difficult because in most cases, library Web site and local digital collection use are not authenticated. Authentication automatically associates transactions with a user session, clearly defining a "visit." In an unauthenticated environment where transactions are associated with IP addresses and public computers are used by many different people, perhaps in rapid succession, defining a visit is not easy.

While the bulk of the discussion centers on what constitutes a visit and how to count the number of visits, one library participating in the DLF survey wants to gather the following data, though it is unclear why this level of specificity was desirable or how the data would be used:

- Number and percentage of Web site visits at time of day and day of week
- Number and percentage of visits that look at one Web page, 2-4 Web pages, 5-10 Web pages, or more than 10 pages
- Number and percentage of visits that last less than 1 minute, 2-4 minutes, 5-10 minutes, or more than 10 minutes per page, service, or collection

However a visit is defined, in an unauthenticated environment the data will be dirty. Libraries are probably prepared to settle for "good-enough" data, but a standard definition would facilitate comparisons across institutions.

Similarly, libraries would like to be able to count e-reserves, e-book, and e-journal use and combine this information with traditional reserves, book, and journal usage statistics to get a complete picture of library use. Again, tracking use of e-resources in a way that is comparable to traditional measures is problematic. Even when e-resources are managed locally, the counts are not comparable, because page hits, not title hits, are logged. Additional work is required to generate hits by title.

In the absence of standards or guidelines, libraries are charting their own course. For example, one site participating in the DLF survey is devising statistics to track use of Web-accessible, low-resolution images, and requests for high-resolution images that are not available on the Web. They are grappling with how to incorporate into their purview metadata from other digital collections available on campus so that they can quantify use of their own content and other campus content. No explanation was offered for how these data would be used.

3.5.1.2. OPAC and Integrated Library Systems

ILS vendors often provide minimal transaction logging because of the high use of the system by staff and end users and the rapid rate with which log files grow to enormous size. When the server is filled with log files, the system ceases to function properly. Many libraries

are not satisfied with the data available for mining in their ILS or the routine reporting mechanisms provided by the vendor. Some libraries have developed custom reports in response to requests from library administrators or department heads. These reports are difficult to produce, often requiring expensive Application Program Interface (API) training from the vendor. Many sites want reports that they cannot produce because they do not have the resources or because the system does not log the information they need. For example, if a library wants to assess market penetration of library books, its ILS might not be able to generate a report of the number of unique users who have checked out books within a specified period of time. If administrators want to determine which books to move to off-site storage, their ILS might not be able to generate a report of which books circulated fewer than five times within a specified period of time.

3.5.1.3. Remote Electronic Resources

Getting the right data from commercial vendors is a well-known problem. Data about use of commercial resources are important to libraries, because use is a measure of service provided and because the high cost of e-resources warrants scrutiny. The data might also be needed to justify subscription expenditures to university administrators. DLF respondents had the usual complaints about vendor-supplied usage statistics:

- The incomparability of the data
- The multiple formats, delivery methods, and schedules for providing the data (for example, e-mail; paper; remote access at the vendor's Web site; monthly, quarterly, annual, or irregular reporting)
- The lack of useful data (for example, no data on use of specific e-resource titles)
- The lack of intelligible or comprehensible data
- The level of specificity of usage data by IP address
- The failure of some vendors to provide usage data at all

While acknowledging that some vendors are collaborating with libraries and making progress in providing useful statistics, libraries continue to struggle to understand what vendors are actually counting and the time periods covered in their reports. Many libraries distrust vendor-supplied data and rue the inability to corroborate these data. One DLF respondent told a story of a vendor calling to report a large number of turn-aways. The vendor encouraged the library to increase the number of licensed simultaneous users. Instead, the library examined the data, noticed the small number of sessions during that two-day period, concluded that the problem was technical, and did not change its license—which was the right course of action. The number of turn-aways was insignificant thereafter. Another story concerned vendor-supplied data about average session lengths. The vendor reported average session lengths of 25 to 26 minutes, but the vendor does not distinguish time-outs from log-outs. Libraries know that many users neglect to log out and that session length is

skewed by users who walk away and the system times out minutes later.

In the absence of standard definitions and staridardized procedures for capturing data about human-computer interactions, libraries cannot compare the results of transaction log analyses across institutions or even across databases and collections within their institutions. Efforts continue to persuade vendors to log standard transactions, extract the data using standard definitions, and provide that information to libraries in standard formats. Meanwhile, libraries remain at the mercy of vendors. Getting meaningful, manageable vendor statistics remains a high priority. Many librarians responsible for licensing e-resources are instructed to discuss usage statistics[6] with vendors before licensing their products. Some librarians are lobbying not to sign contracts if the vendor does not provide good statistics. Nevertheless, vendors know that useful statistics are not yet required to make the sale.

3.5.2. Analyzing and Interpreting the Data

DLF respondents understand that usage statistics are an important measure of library service and, to some degree, an indication of user satisfaction. Usage data must be interpreted cautiously, however, for two reasons. First, usability and user awareness affect the use of library collections and services. Low use can occur because the product's user interface is difficult to use, because users are unaware that the product is available, or because the product does not meet the users' information needs. Second, usage statistics do not reveal the users' experience or perception of the utility or value of a collection or service. For example, though a database or Web page is seldom used, it could be very valuable to those who use it. The bottom line is that usage statistics provide necessary but insufficient data to make strategic decisions. Additional information, gathered from user studies, is required to provide a context in which to interpret usage data.

Many DLF respondents observed that reports generated by TLA are not analyzed and applied. Perhaps this is because the library lacks the resources or skills to do the work. It may also be because the data lack context and interpretation is difficult. Several respondents requested guidance in how to analyze and interpret usage data and diagnose problems, particularly with use of the library Web site.

3.5.3. Managing, Presenting, and Using the Data

DLF libraries reported needing assistance with how to train their staff to use the results of the data analysis. The problem appears to be exacerbated in decentralized library systems and related to the difficulty of compiling and manipulating the sheer bulk of data generated by TLA. Monthly reports of Web site use, digital collection use, and remote e-resource use provide an overwhelming volume of

[6] To guide these discussions, libraries are using the International Coalition of Library Consortia (ICOLC) Guidelines for Statistical Measures of Usage of Web-Based Indexed, Abstracted, and Full-Text Resources. Available at: http://www.library.yale.edu/consortia/Webstats.html.

information. Libraries expressed concern that they were not taking full advantage of the information they collect because they do not have the resources to compile it. Vendor statistics are a well-known case in point.

Because of the problems with vendor statistics, management and analysis of the data are cumbersome, tedious, and time-consuming. If the data are compiled in any way, typically only searches, sessions, and full-text use are included for analysis. Some DLF libraries gather and compile statistics from all vendors. Some compile usage statistics only on full-text journals and selected large databases. Some compare data only within products provided by a single vendor, not across products provided by different vendors. Others use data from different vendors to make comparisons that they know are less than perfect, or they try to normalize the data from different vendors to enable cross-product comparisons. For example, one site uses the number of sessions reported by a vendor to predict the number of searches of that vendor's product based on the ratio of searches to sessions from comparable e-resources. Libraries that compile vendor statistics for staff or consortium perusal provide access to the data using either a spreadsheet or an IP-address-restricted Web page. One site described the painstaking process of producing this Web page: entering data from different vendor reports—from e-mail messages, printed reports, downloaded statistics—into a spreadsheet, then using the spreadsheet to generate graphs and an HTML table for the Web. The time and cost of this activity must be weighed against the benefits of such compilations.

Even if e-resource usage data are compiled, libraries struggle with how to organize and present the information to an audience for consideration in decision making and strategic planning. For example, how should monthly usage reports of 800 e-journals be organized? The quality of the presentation can affect the decisions made based on the data. Training is required to make meaningful, persuasive graphical presentations. Libraries need guidance in how to manage, present, and apply usage data effectively.

4. GENERAL ISSUES AND CHALLENGES

4.1. Issues in Planning a Research Project

When a decision to conduct research has been made, a multifaceted process begins. Each step of that process requires different knowledge and skills. Whatever the research method, all research has certain similarities. These relate to focusing the research purpose, marshalling the needed resources, and scheduling and assigning responsibilities. Conducting user studies also requires selecting a sampling method, recruiting subjects, and getting approval from the IRB to conduct research with human subjects.

The experiences reported by DLF respondents underscore the importance of careful planning and a comprehensive understanding of the full scope of the research process. Textbooks outline the plan-

ning process. It begins with articulating the research purpose. The second step is conducting an assessment of human and financial resources available to conduct the research and clearly assigning who is responsible for each stage of the process—designing the research instruments; preparing the schedule; gathering, analyzing, and interpreting the data; presenting the findings; and developing and implementing plans to use them. The third step is selecting the research method (Chadwick, Bahr, and Albrecht 1984). The frequent breakdowns that DLF libraries experience in the research process suggest problems in planning, particularly in marshalling the resources needed to complete the project. Perhaps those responsible for planning a study do not have enough power or authority to assemble the requisite human and financial resources. Perhaps they do not have the time, resources, or understanding of the research process to develop a comprehensive plan. Whatever the case, resources assigned to complete research projects are often insufficient. The breakdown often occurs at the point of developing and implementing plans to use the research results. The process of developing a plan can get bogged down when the results are difficult to interpret. Implementing plans can get bogged down when plans arrive on the doorstep of programmers or Web masters who had no idea the research would create work for them. Data can go unused if commitment has not been secured from every unit and person necessary to complete a project. Even if commitment is secured during the planning stage, if a project falls significantly behind schedule, other projects and priorities can intervene, and the human resources needed to implement research results will not be available when they are needed.

Scheduling also influences the success or failure of research efforts. Many DLF respondents reported underestimating the time it takes to accomplish different steps in the research process. Getting IRB approval to conduct human subjects research can take months. Recruiting research subjects can be time-consuming. Analyzing and interpreting the data and documenting the research findings can take as much time as planning the project, designing the research instruments and procedures, and gathering the data. The time it takes to implement a plan depends on the plan itself and competing priorities of the implementers. An unrealistic schedule can threaten the success of the project. A carefully constructed schedule can facilitate effective allocation of resources and increase the likelihood that research results will be applied. Comments from DLF respondents suggest that the larger the number of persons involved in any step of this process, the longer the process takes. Cumbersome governance of user studies can be counter-productive.

The limitations of research results and the iterative nature of the research process also challenge DLF libraries. Additional research is often necessary to interpret survey data or to identify solutions to problems that surface in user protocols. Realizing that multiple studies might be necessary before concrete plans can be formulated and implemented can be discouraging. Conducting research can seem like an endless loop of methods and studies designed to identify

problems, determine how to solve them, and verify that they have been solved. When a library's resources are limited, it is tempting to go with intuition or preferences. Nevertheless, DLF respondents agreed that libraries must stay focused on users. Assessment must be an ongoing priority. Research must be iterative, because user needs and priorities change with time and technology. To provide quality service, the digital library must keep pace with users.

Multiple research methods and a sequence of studies are required for the digital library to evolve in a way that serves users well. DLF respondents reported the following cases, which illustrate the rich, although imperfect, benefits that derive from triangulated or iterative efforts.

- **Protocol, Transaction Log, and Systems Analysis Research.** Think-aloud user protocols were conducted in a laboratory to assess the usability of the library Web site. The study focused on the home page and e-resources and databases pages. A task script was prepared in consultation with a commercial firm. Its purpose was to identify the 10 tasks most frequently performed by students, faculty, and staff on the library's Web site. Another firm was hired to analyze the Web site architecture, transaction logs, and usability (protocol) data and to conduct additional research to capture user perceptions of the Web site. On the basis of these analyses, the firm provided an interface design specification, architectural framework, and short- and long-term goals for the Web site. The firm also recommended the staffing needed to maintain the proposed architecture. The library used the design specification to revise its Web site, but the recommendations about staffing to maintain the Web site did not fit the political environment of the library. For example, the recommendation included creating an advisory board to make decisions about the Web site, hiring a Web master, and forming a Web working group to plan Web site development. The library has a Web working group and has created a new Web coordinator position, but is having trouble filling it. Librarians believe the issue is lack of ownership of Web project management. No advisory board was created.
- **Heuristic Evaluation, Card Sorting, Protocol, and Survey Research.** A library created a task force to redesign the library Web site on the basis of anecdotal evidence of significant problems and the desire for a "fresh" interface. The task force
  - Conducted a heuristic evaluation of the existing library Web site
  - Looked at other Web sites to find sites its members liked
  - Created a profile of different user types (for example, new or novice users, disabled users)
  - Created a list of what the redesigned Web site had to do, organized by priority
  - Created a content list of the current Web site that revealed content of interest only to librarians (for example, a list of library organizations)

- Created a content list for the redesigned Web site that eliminated any content in the existing site that did not fit the user profiles
- Conducted a card-sorting study to help group items on the content list
- Conducted a Web-based survey to help determine the vocabulary for group and item (link) labels. (The survey did not work very well because the groups and items the participants were to label were difficult to describe.)
- Implemented a prototype of the new library Web site home page and secondary pages
- Conducted think-aloud protocols with the prototype Web pages. (The library recruited and screened participants to get eight subjects. The subjects signed consent forms, then did the protocol tasks. Different task scripts were provided for undergraduate students, graduate students, and faculty. The protocols were audiotaped and capture software was used to log participant keystrokes. The facilitator also took notes during the protocols. The results of the protocol study revealed that many of the problems users encountered were not user interface problems, but bibliographic instruction problems.)
- Conducted a survey questionnaire to capture additional information about the participants' experience and perception of the new Web site

Although these activities took a substantial amount of time, they were easy and inexpensive to do and were very revealing. The new Web sites were a significant improvement over the old sites. User studies will be conducted periodically to refine the design and functionality of the sites.

The purpose of the usability studies and many of the other user studies described in this report is to improve interface design and functionality. One experienced DLF respondent outlined the following as the ideal, iterative process to implement a user-friendly, fully functional interface:

1. Develop a paper prototype in consultation with an interface design expert applying heuristic principles of good design.
2. Conduct paper prototype and scenario research.
3. Revise the paper prototype on the basis of user feedback and heuristic principles of good design.
4. Conduct paper prototype and scenario research.
5. Revise the design on the basis of user feedback and implement a functioning prototype.
6. Conduct think-aloud protocols to test the functionality and navigation of the prototype.
7. Revise the prototype on the basis of user feedback and heuristic principles of good design.
8. Conduct think-aloud protocols to test the new design.
9. Revise the design on the basis of user feedback.
10. Release the product.

11. Revise the design on the basis of user feedback and analysis of transaction logs.

Libraries would benefit greatly from sharing their experiences and developing guidelines for planning and scheduling different kinds of studies and iterations. An outline of the key decision points and pitfalls would be an ideal way to share lessons learned. Similarly, libraries would benefit from discussing and formulating a way to integrate assessment into the daily fabric of library operations, to make it routine rather than remarkable, and thereby possibly avoid generating unnecessary and unhelpful comments and participation.

4.2. Issues in Implementing a Research Project

Several issues in implementing a research project have already been described. For example

- Selecting the appropriate research method for the research purpose
- Developing effective and appropriate research instruments
- Developing the requisite skills to conduct research using different methods, including how to gather, analyze, interpret, and present the data effectively, and how to develop plans
- Developing a system or method to manage data over time
- Organizing assessment as a core activity
- Allocating sufficient human and financial resources to conduct and apply the results of different research methods
- Developing comprehensive plans and realistic schedules to conduct and apply the results of different research methods (the academic calendar affects the number of participants who can be recruited and when the results can be applied)
- Maintaining focus on users when research results challenge the operating assumptions and personal preferences of librarians
- Recruiting representative research subjects who meet the criteria for the study (for example, subjects who can think aloud, subjects experienced or not experienced with the product or service being studied)

DLF respondents discussed two additional issues that affect user studies: sampling and getting IRB approval to conduct human subjects research. Sampling is related to the problem of recruiting representative research subjects. IRB approval relates to planning and scheduling research and preserving the anonymity of research subjects.

4.2.1. Issues in Sampling and Recruiting Research Subjects

Sampling is the targeting and selection of research subjects within a larger population. Samples are selected on the basis of the research purpose, the degree of generalization desired, and available resources. The sample ideally represents the entire target population. To be representative, the sample must have the characteristics of the target population, preferably in the proportion they are found in the larger

population. To facilitate selecting representative samples, sampling units or groups are defined within a population. For example, in a university, the sampling units are often undergraduate students, graduate students, and faculty. Depending on the purpose of the study, the sampling units for a study of undergraduate students could be based on the school or college attended (for example, fine arts, engineering) or the class year (for example, freshmen/sophomore, junior/senior). Though research typically preserves the anonymity of research subjects, demographic data are captured to indicate the sampling unit and other nonidentifying characteristics of the participants considered relevant to the study (for example, faculty, School of Business).

Textbooks outline several different methods for selecting subjects from each sampling unit designated in a study:

- **Random sampling**. To represent the target population accurately, a sample must be selected following a set of scientific rules. The process of selecting research subjects at random, where everyone in the target population has the same probability of being selected, is called random sampling. There are many methods for random sampling units within a larger population. Readers are advised to consult an expert or a textbook for instruction.
- **Quota sampling**. Quota sampling is the process of using information about selected characteristics of the target population to select a sample. At its best, quota sampling selects a sample with the same proportion of individuals with these characteristics as exists in the population being studied. How well quota samples represent the target population and the accuracy of generalizations from quota sample studies depends on the accuracy of the information about the population used to establish the quota.
- **Convenience sampling**. The process of selecting research subjects and sampling units that are conveniently available to the researcher is called convenience sampling. The results of studies conducted with convenience samples cannot be generalized to a larger population because the sample does not represent any defined population.

Two additional sampling methods might produce a representative sample, but there is no way to verify that the sample actually represents the characteristics of the target population without conducting a study of a representative (random) sample of the population and comparing its characteristics with those of the sample used in the initial study. These methods are as follows:

- **Purposive sampling**. This activity entails selecting research subjects and sampling units on the basis of the expertise of the researcher to select representatives of the target populations.
- **Snowball sampling**. This process entails identifying a few research subjects who have the characteristics of the target population and asking them to name others with the relevant characteristics.

DLF libraries have used all of these sampling methods to select human subjects for user studies. For example, a library conducted a survey to assess journal collection use and need by mailing a survey to a statistically valid, random sample of faculty and graduate students. It used the characteristics of reference service users to target and select the sample for a survey about reference service. In rare cases, all the users of a service have been invited to participate in a study (for example, all the graduate students and faculty with assigned study carrels). In many cases, however, libraries conduct user studies with convenience samples that fall short of accurately representing the sampling units within the target population. Sometimes librarians provide the names of potential research subjects, which can skew the data toward experienced users.

Recruiting research subjects is so time consuming that the emerging practice is to provide financial or other incentives to recruit enough volunteers to "take the temperature" of what is going on with users of particular library services, collections, or interfaces. Though providing incentives can bias the research results, many DLF respondents commented that some user feedback is better than none. Libraries are experimenting with providing different incentives. With surveys, the names of participants are gathered (apart from the survey data, to ensure anonymity), and one or more names are drawn to win cash or some other prize. Every student in a focus group or think-aloud protocol study might be given $10 or $20 or a gift certificate to the bookstore, library coffee shop, or local movie theatre. Often lunch is provided to recruit students or faculty to participate in focus groups. Some libraries are considering providing more substantial rewards, such as free photocopying. Recruiting faculty can be particularly difficult because the incentives that libraries can afford to offer are inadequate to get their interest. Holding a reception during which the research results are presented and discussed is one way to capture faculty participation.

DLF libraries prefer to have hundreds of people complete formal survey questionnaires, with respondents ideally distributed in close proportion to the representation of sampling units on campus. They conduct focus groups with as few as six subjects per sampling unit, but prefer eight to ten participants per group. Many DLF respondents were comfortable with Nielsen's guideline of using four to six participants per sampling unit in think-aloud protocol studies. A few questioned the validity of Nielsen's claims, referencing the "substantial debate" at the Computer-Human Interaction 2000 Conference about whether some information was better than none. Others questioned whether six to eight subjects are enough in a usability study in the library environment, where users come from diverse cultural backgrounds. Given the work being done on such things as how cultural attitudes toward technology and cultural perceptions of interpersonal space affect interface design and computer-mediated communication,[7] how does or should diversity affect the design of digital library collections and services?

[7] See, for example, Ess 2001.

Lack of a representative sample raises questions about the reliability and validity of data, particularly when studies are conducted with small samples and few sampling units. Using finer-grain sampling units and recruiting more subjects can increase the degree to which the sample is representative and address concerns about diversity. For example, instead of conducting one focus group with undergraduate students, a library could conduct a focus group with undergraduate students in each school or college in the university or a focus group with undergraduates from different cultural backgrounds—Asian, African-American, and Hispanic. The disadvantage of this approach is that it will increase the cost of the research.

DLF respondents indicated that they were willing to settle for a "good-enough" distribution of user groups, but were wrestling with how to determine and recruit a "good-enough" sample. There is inevitably a trade-off between the cost of recruiting additional research subjects and its benefits. Finding the appropriate balance seems to hinge on the goal of the assessment. Accuracy and the costs associated with it are essential in a rigorous experiment designed to garner precise data and predicative results, but are probably not essential when the goal is to garner data indicative and suggestive of trends. Focusing on the goal of identifying trends to help shape or improve user service could assuage much of the angst that librarians feel about the validity of their samples and the results of their research.

4.2.2. Issues in Getting Approval and Preserving Anonymity

Research must respect the dignity, privacy, rights, and welfare of human beings. Universities and other institutions that receive funding from federal agencies have IRBs that are responsible for ensuring that research will not harm human subjects, that the subjects have given informed consent, and that they know they may ask questions about the research or discontinue participating in it at any time. In providing informed consent, research subjects indicate that they understand the nature of the research and any risks to which they will be exposed by participating, and that they have decided to participate without force, fraud, deceit, or any other form of constraint or coercion.

DLF respondents were aware of IRB requirements. Some expressed frustration with their IRB's turn-around time and rules. Others had negotiated blanket approval for the library to conduct surveys, focus groups, and protocols and therefore did not need to allow time to get IRB approval for each study.

To apply for IRB approval, libraries must provide the IRB with a copy of the consent form that participants will be required to read and sign, and a brief description of the following:

- Research method
- Purpose of the research
- Potential risks and benefits to the research subjects
- How the privacy and anonymity of research subjects will be preserved
- How the data will be analyzed and applied

- How, where, and for how long the data will be stored
- Who will conduct the research
- Who will have access to the data

On grant-funded projects, the signatures of the principal investigators are required on the application for IRB approval, regardless of whether they themselves will be conducting the human subjects research. A recent development requires completion of an online tutorial on human subjects research that culminates in certification. The certificate must be printed and submitted to the IRB.

Typically, IRB approval to conduct a particular study is granted for one year. If the year ends before the research with human subjects is completed, the researcher must follow the same procedures to apply for renewal. If the IRB does not grant blanket approval to conduct particular kinds of research, whether DLF libraries seek IRB approval for all user studies or just those funded by the federal government is a matter of local policy.

IRB guidelines, regulations, and other documents are available at the Web site of the Office for Human Research Protections, U.S. Department of Health and Human Services at http://ohrp.osophs.dhhs.gov/.

No DLF respondent addressed whether IRB approval was secured for routine transaction logging of use of its Web site, OPAC, ILS, proxy server, or local digital collections. Several respondents did indicate, however, that they are uncertain whether users know that they are tracking these transactions. The issue is of some concern with authenticated access because it identifies individual users. If authentication data are logged, they can be used to reconstruct an individual's use of the digital library. Even if the data are encrypted, the encryption algorithm can be compromised. Few libraries require users to authenticate before they can use public computers in the library, and access to remote electronic resources is typically restricted by IP address. However, authentication is required for all proxy server users and users with personalized library Web pages. Many, or most, libraries run a proxy server, and personalized Web pages are growing in popularity. Personalized Web pages enable libraries to track who has what e-resources on their Web pages and when they use these resources. Authentication data in proxy server logs can be used to reconstruct individual user behavior. Card-swipe exit data also identify individuals and can be used to reconstruct the date, time, and library they visited. The adoption of digital certificates will enable the identification and tracking of an individual's use of any resource that employs the technology.

While library circulation systems have always tracked the identity of patrons who borrow traditional library materials, the association between the individual and the items is deleted when the materials are returned. Government subpoenas could force libraries to reveal the items that a patron currently has checked out, but the library does not retain the data that would be required to reveal a patron's complete borrowing history. In the case of transaction logs,

however, the association remains as long as the library maintains the log files, unless the library manipulates the files in some way (for example, by replacing individual user IDs with the school and status of the users). Without such manipulation, it is possible for libraries, hackers, or government agencies to track an individual's use of digital library collections and services over whatever period of time the log files are maintained. While there could be good reason to track the usage patterns of randomly selected individuals throughout their years at the university, the possibility raises questions about informed consent and perhaps challenges the core value of privacy in librarianship. The effects of the recently passed Anti-Terrorism Act on the privacy of library use are not yet known.

5. CONCLUSIONS AND FUTURE DIRECTIONS

Libraries face five key challenges related to assessment:

1. Gathering meaningful, purposeful, comparable data
2. Acquiring methodological guidance and the requisite skills to plan and conduct assessments
3. Managing assessment data
4. Organizing assessment as a core activity
5. Interpreting library trend data in the larger environmental context of user behaviors and constraints

Libraries urgently need statistics and performance measures appropriate to assessing traditional and digital collections and services. They need a way to identify unauthenticated visits to Web sites and digital collections, as well as clear definitions and instructions for compiling composite input and output measures for the hybrid library. They need guidelines for conducting cost-effectiveness and cost-benefit analyses and benchmarks for making decisions. They need instruments to assess whether students are really learning by using the resources libraries provide. They need reliable, comparative, quantitative baseline data across disciplines and institutions as a context for interpreting qualitative and quantitative data indicative of what is happening locally. They need assessments of significant environmental factors that may be influencing library use in order to interpret trend data. To facilitate comparative assessments of resources provided by the library, by commercial vendors, and by other information service providers, DLF respondents commented that they need a central reporting mechanism, standard definitions, and national guidelines that have been developed and tested by librarians, not by university administrators or representatives of accreditation or other outside agencies.

Aggressive efforts are under way to satisfy all of these needs. For example, the International Coalition of Library Consortia's (ICOLC) work to standardize vendor-supplied data is making headway. The Association of Research Libraries' (ARL) E-metrics and LIBQUAL+ efforts are standardizing new statistics, performance measures, and

research instruments. Collaboration with other national organizations, including the National Center for Education Statistics (NCES) and the National Information Standards Organization (NISO), shows promise for coordinating standardized measures across all types of libraries. ARL's foray into assessing costs and learning and research outcomes could provide standards, tools, and guidelines for these much-needed activities as well. Their plans to expand LIBQUAL+ to assess digital library service quality and to link digital library measures to institutional goals and objectives are likely to further enhance standardization, instrumentation, and understanding of library performance in relation to institutional outcomes. ARL serves as the central reporting mechanism and generator of publicly available trend data for large research libraries. A similar mechanism is needed to compile new measures and disseminate trend data for other library cohort groups.

Meanwhile, libraries have diverse assessment practices and sometimes experience failure or only partial success in their assessment efforts. Some DLF respondents expressed dismay at the pace of progress in the development of new measures. The pace is slower than libraries might like, in the context of the urgency of their need, because developing and standardizing assessment of current library resources, resource use, and performance is very difficult. Libraries are in transition. It is hard to define, let alone standardize, what libraries do, or to measure how much they do or how well they do it, because what they do is constantly changing. Deciding what data to collect and how to collect them are difficult because library collections and services are evolving rapidly. New media and methods of delivery evolve at the pace of technological change, which, according to Raymond Kurzweil (2000), doubles every decade.[8] The methods for assessing new resource delivery evolve at a slower rate than do the resources themselves. This is the essential challenge and rationale for the efforts of ARL, ICOLC, and other organizations to design and standardize appropriate new measures for digital libraries. It also explains the difficulties involved in developing good trend data and comparative measures. Even if all libraries adopted new measures as soon as they became available, comparing the data would be difficult because libraries evolve on different paths and at different rates, and offer different services or venues for service. Given the context of rapid, constant change and diversity, the new measures initiatives are essential and commendable. Without efforts on a national scale to develop and field test new measures and build a consensus, libraries would hesitate to invest in new measures. Just as absence of community agreement about digitization and metadata standards is an impediment to libraries that would otherwise digitize some of their collections, lack of community agreement about appropriate new measures is an impediment to investing in assessment.

[8] Kurzweil is founder and chief technology officer, Kurzweil Applied Intelligence, and founder and chief executive officer, Kurzweil Educational Systems.

Despite the difficulties, substantial progress is being made. Consensus is being achieved. Libraries are slowly adopting composite measures, such as those developed by John Carlo Bertot, Charles McClure, and Joe Ryan, to capture traditional and digital library inputs, outputs, and performance. For example[9]

- Total library visits = total gate counts + total virtual visits
- Percentage of total library visits that are virtual
- Total library materials use = total circulation + total in-house use of materials + total full-text electronic resources viewed or downloaded
- Percentage of total library materials used in electronic format
- Total reference activity = total in-person transactions + total telephone transactions + total virtual (for example, e-mail, chat) transactions
- Percentage of total reference activity conducted in virtual format
- Total serials collection = total print journal titles + total e-journal titles
- Percentage of total serials collection available in electronic format

Analysis of composite measures over time will provide a more comprehensive picture of what is happening in libraries and will enable libraries to present more persuasive cases to university administrators and other funders to support libraries and their digital initiatives. Perhaps a lesson learned in system development applies here. Interoperability is possible when a limited subset of metadata tags and service offerings are supported. In the context of assessment, a limited subset of statistics and performance measures could facilitate comparison yet also allow for local variations and investments. ARL is taking this approach in its effort to develop a small set of core statistics for vendor products.

Reaching a consensus on even a minimum common denominator set of new statistics and performance measures would be a big step forward, but libraries also need methodological guidance and training in the requisite skills. Practical manuals and workshops, developed by libraries for libraries, that describe how to gather, analyze, interpret, present, and apply data to decision making and strategic planning would facilitate assessment and increase return on the investment in assessment. ARL is producing such a manual for E-metrics. The manual will provide the definition of each measure, its rationale, and instructions for how to collect the data. ARL also offers workshops, Systems and Procedures Exchange Center (SPEC) kits, and publications that facilitate skill development and provide models for gathering, analyzing, and interpreting data. However, even if libraries take advantage of ARL's current and forthcoming offerings, comments from DLF respondents indicate that gaps remain in several areas.

---

[9] The measures were developed for public library network services, but are equally suited to academic libraries. See *Statistics and Performance Measures for Public Library Network Services*. 2000. Chicago: American Library Association.

"How-to" manuals and workshops are greatly needed in the area of user studies. Although DLF libraries are conducting a number of user studies, many respondents asked for assistance. Manuals and workshops developed by libraries for libraries that cover the popular assessment methods (surveys, focus groups, and user protocols) and the less well-known but powerful and cost-effective discount usability testing methods (heuristic evaluations and paper prototypes and scenarios) would go a long way toward providing such guidance. A helpful manual or workshop would

- Define the method
- Describe its advantages and disadvantages
- Provide instruction in how to develop the research instruments and gather and analyze the data
- Include sample research instruments proven successful in field testing
- Include sample quantitative and qualitative results, along with how they were interpreted, presented, and applied to realistic library concerns
- Include sample budgets, time lines, and workflows

Standard, field-tested research instruments for such things as OPAC user protocols or focus groups to determine priority features and functionality for digital image collections would enable comparisons across libraries and avoid the cost of duplicated efforts in developing and testing the instruments. Similarly, budgets, time lines, and workflows derived from real experience would reduce the cost of trial-and-error efforts replicated at each institution.

The results of the DLF study also indicate that libraries would benefit from manuals and workshops that provide instruction in the entire research process—from conception through implementation of the results—particularly if attention were drawn to key decision points, potential pitfalls, and the skills needed at each step of the process. Recommended procedures and tools for analyzing, interpreting, and presenting quantitative and qualitative data would be helpful, as would guidance in how to turn research findings into action plans. Many libraries have already learned a great deal through trial and error and through investments in training and professional development. Synthesizing and packaging their knowledge and expertise in the form of guidelines or best practices and disseminating it to the broader library community could go a long way toward removing impediments to conducting user studies and would increase the yield of studies conducted.

TLA presents a slightly different set of issues because the data are not all under the control of the library. Through the efforts of ICOLC and ARL, progress is being made on standardizing the data points to be delivered by vendors of database resources. ARL's forthcoming instruction manual on E-metrics will address procedures for handling these vendor statistics. Similar work remains to be done with OPAC and ILS vendors and vendors of full-text digital collections. Library-managed usage statistics for their Web sites and local

databases and digital collections present a third source of TLA data. Use of different TLA software, uncertainty or discrepancy in how the data points are defined and counted, and needed analyses not supported by some of the software complicate data gathering and comparative analysis of use of these different resources. Work must be done to coordinate efforts on all these fronts to facilitate comparative assessments of resources provided by the library, commercial vendors, and other information service providers.

In the meantime, libraries could benefit from guidance on how to compile, interpret, present, and use the TLA data they do have. For example, DLF libraries have taken different approaches to compiling and presenting vendor data. A study of these approaches and the costs and benefits of each approach would be instructive. Case studies of additional research conducted to provide a context for interpreting and using TLA data would likewise be informative. For example, what does the increasing or decreasing number of queries of licensed databases mean? Is an increase necessarily a good thing and a decrease necessarily a bad thing? Does a decrease indicate a poor financial investment? Could a decrease in the number of queries simply mean that users have become better searchers? What do low-use or no-use Web pages mean? Poor Web site design? Or wasted resources producing pages of information that no one needs? Libraries would benefit if those who have gathered data to help answer these questions would share what they have learned.

The issue of compiling assessment data is related to managing the data and generating trend lines over time. Libraries need a simplified way to record and analyze input and output data on traditional and digital collections and services, as well as an easy way to generate statistical reports and trend lines. Several DLF libraries reported conducting needs assessments for library statistics in their institutions, eliminating data-gathering practices that did not address strategic concerns or were not required for internal or external audiences. They also mentioned plans to develop a homegrown MIS that supports the data manipulations they want to perform and provides the tools to generate the graphics they want to present. Designing and developing an MIS could take years, not counting the effort required to train staff how to use the system and secure their commitment to use it. Only time will tell whether the benefits to individual libraries will exceed the cost of creating these homegrown systems.

The fact that multiple libraries are engaged in this activity suggests a serious common need. One wonders why a commercial library automation vendor has not yet marketed a product that manages, analyzes, and graphically presents library data. The local costs of gathering, compiling, analyzing, managing, and presenting quantitative data in effective ways, not to mention the cost of training and professional development required to accomplish these tasks, could exceed the cost of purchasing a commercial library data management system, were such a system available. The market for such a system would probably be large enough that a vendor savvy enough to make it affordable could also make it profitable. Such a system

would reduce the need for librarians to interpret and apply data effectively. The cost savings would be spent on purchasing the system. The specifications and experiences of libraries engaged in creating their own MIS could be used to develop specifications for the design of a commercial MIS. Building a consensus within the profession for the specification and marketing it to library automation vendors could yield collaborative development of a useful, affordable system. Admittedly, the success of such a system depends in part on the entry and verification of correct data, but this issue could begin to resolve itself, given standard data points and a system, designed by libraries for libraries, that saves resources and contributes to strategic planning.

The results of the DLF study suggest that individually, libraries in many cases are collecting data without really having the will, organizational capacity, or interest to interpret and use the data effectively in library planning. Libraries have been slow to standardize definitions and assessment methods, develop guidelines and best practices, and provide the benchmarks necessary to compare the results of assessments across institutions. These problems are no doubt related to the fact that library use and library roles are in continuous transition. The development of skills and methods cannot keep pace with the changing environment. The problems may also be related to the internal organization of libraries. Comments from DLF respondents indicate that the internal organization of many libraries does not facilitate the gathering, analysis, management, and strategic use of assessment data. The result is a kind of purposeless data collection that has little hope of serving as a foundation for the development of guidelines, best practices, or benchmarks. The profession could benefit from case studies of those libraries that have conducted research efficiently and applied the results effectively. Understanding how these institutions created a program of assessment—how they integrated assessment into daily library operations, how they organized the effort, how they secured commitment of human and financial resources, and what human and financial resources they committed—would be helpful to the many libraries currently taking an ad hoc approach to assessment and struggling to organize their effort. Including budgets and workflows for the assessment program would enhance the utility of such case studies.

Efforts to enhance research skills, to conduct and use the results of assessments, to compile and manage assessment data, and to organize assessment as a core library activity all shed light on *how* libraries and library use are changing. What remains to be known is *why* libraries and library use are changing. To date, speculation and intuition have been employed to interpret known trends; however careful interpretation of the data requires knowledge of the larger context within which libraries operate. Many DLF respondents expressed a need to know what information students and faculty use, why they use this information, and what they do or want to do when they need information or when they find information. Respondents acknowledged that these behaviors, including use of the li-

brary, are constrained by changes on and beyond the campus, including the following:

- Changes in the habits, needs, and preferences of users; for example, undergraduate students now turn to a Web search engine instead of the library when they need information
- Changes in the curriculum; for example, elimination of research papers or other assignments that require library use, distance education courses, or the use of course packs and course management software that bundle materials that might otherwise have been found in the library
- Changes in the technological infrastructure; for example, penetration and ownership of personal networked computers, network bandwidth, or wireless capabilities on university and college campuses that enable users to enter the networked world of information without going through pathways established by the library.
- Use of competing information service providers; for example, Ask-A services, Questia, Web sites such as LibrarySpot, or the Web in general

In response to this widespread need to know, the Digital Library Federation, selected library directors, and Outsell, Inc., have designed a study to examine the information-seeking and usage behaviors of academic users. The study will survey several thousand students and faculty in different disciplines and different types of institutions to begin to understand how they perceive and use the broader information landscape. The study will provide a framework for understanding how academics find and use information (regardless of whether the information is provided by libraries), examine changing patterns of use in relation to changing environmental factors, identify gaps where user needs are not being met, and develop baseline and trend data to help libraries with strategic planning and resource allocation. The findings will help libraries focus their efforts on current and emerging needs and expectations of academic users, evaluate their current position in the information landscape, and plan their future collections, services, and roles on campus on the basis of an informed, rather than speculative, understanding of academic users and uses of information.[10]

The next steps recommended based on the results of the DLF study are the collaborative production and dissemination of the following:

- E-metrics lite: a limited subset of digital library statistics and performance measures to facilitate gathering baseline data and enable comparisons
- How-to manuals and workshops for
  - conducting research in general, with special emphasis on planning and commitment of resources

[10] The research proposal and plans are available at http://www.diglib.org/use/grantpub.pdf.

    - conducting and using the results of surveys, focus groups, user protocols, and discount usability studies, with special emphasis on field-tested instruments, time lines, budgets, workflows, and requisite skills
- Case studies of
    - the costs and benefits of different approaches to compiling, presenting, interpreting, and using vendor TLA data in strategic planning
    - how institutions successfully organized assessment as a core library activity
    - a specification for the design and functionality of an MIS to capture traditional and digital library data and generate composite measures, trend data, and effective graphical presentations

Libraries today are clearly needy. Facing rampant need and rapid change, their ingenuity and diligence are remarkable. Where no path has been charted, they carve a course. Where no light shines, they strike a match. They articulate what they need to serve users and their institutional mission, and if no one provides what they need, they provide it themselves, ad hoc perhaps, but for the most part functional. In search of high quality, they know when to settle for good enough—good-enough data, good-enough research and sampling methods, good enough to be cost-effective, good enough to be beneficial to users. In the absence of standards, guidelines, benchmarks, and adequate budgets, libraries work to uphold the core values of personal service and equitable access in the digital environment. Collaboration and dissemination may be the keys to current and future success.

APPENDIX A

References and Selected Bibliography

Web addresses in this appendix were valid as of January 7, 2002.

REFERENCES

Chadwick, B.A., H.M. Bahr, and S.L. Albrecht. 1984. *Social Science Research Methods*. Upper Saddle River, N.J.: Prentice-Hall.

Dillman, Don A. 1978. *Mail and Telephone Surveys: The Total Design Method*. New York: John Wiley.

Ess, Charles, ed. 2001. *Culture, Technology, Communication: Towards an Intercultural Global Village*. New York: SUNY Press.

Greenstein, Daniel and Denise Troll. 2000. Usage, Usability and User Support. Report of a discussion group convened at the DLF Forum on 2 April 2000. Version 1.1. Available at: http://www.diglib.org/use/useframe.htm

Kurzweil, Raymond. 2000. "Promise and Peril: Deeply Intertwined Roles of 21st Century Technology," presentation at the conference "earthware: a good world in 2050 . . . will computers help or hinder?" Carnegie Mellon University, Pittsburgh, Pa., October 19, 2000.

Marczak, Mary, and Meg Sewell. [No date.] Using Focus Groups for Evaluation. *CYFERNet Evaluation*. Available at: http://Ag.Arizona.Edu/fcr/fs/cyfar/focus.htm.

Nielsen, Jakob. 2000. Why You Only Need to Test With 5 Users. *Alertbox*. (March 19). Available at: http://www.useit.com/alertbox/20000319.html.

Nielsen, Jakob. 1994. Heuristic Evaluation. In *Usability Inspection Methods*, edited by Jakob Nielsen and Robert L. Mack. New York: John Wiley and Sons.

Nielsen, Jakob. [No date.] Ten Usability Heuristics. Available at: http://www.useit.com/papers/heuristic/heuristic_list.html.

SELECTED BIBLIOGRAPHY

Social Science Research Methods

Bernard, H. Russell. 2000. *Social Research Methods: Qualitative and Quantitative Approaches.* Thousand Oaks, Calif.: Sage Publications.

Bundy, Mary Lee. 1959. Research Methods in the Social Sciences: A Selected Bibliography Prepared for Students of Library Science. Urbana, Ill.

Caverly, Grant, and Leslie R. Cohen. 1997. *Social Science Research Methods: The Basics.* Greenfield Park, Quebec: C & C Educational Publications.

Chadwick, B.A., H.M. Bahr, and S.L. Albrecht. 1984. *Social Science Research Methods.* Upper Saddle River, N.J.: Prentice-Hall, Inc.

Kenny, Richard F., and David R. Krathwohl. 1993. *Instructor's Manual to Accompany Methods of Educational and Social Science Research: An Integrated Approach.* White Plains, N.Y.: Longman.

King, G., R.O. Keohane, and S. Verba, S. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research.* Princeton, N.J.: Princeton University Press.

Magnani, Robert. Sampling Guide. *Social Science Information Gateway.* Available at: http://www.fantaproject.org/downloads/pdfs/sampling.pdf.

Morgan, G., ed. 1983 *Beyond Method: Strategies for Social Research.* Beverly Hills, Calif.: Sage Publications.

Social Science Information Gateway SOSIG. Qualitative Methods. Available at: http://sosig.esrc.bris.ac.uk/roads/subject-listing/World-cat/qualmeth.html.

Social Science Information Gateway SOSIG. Quantitative Methods. Available at: http://sosig.esrc.bris.ac.uk/roads/subject-listing/World-cat/quanmeth.html.

Summerhill, W.R., and C.L. Taylor. 1992. Selecting a Data Collection Technique. Circular PE-21, Program Evaluation and Organizational Development, Florida Cooperative Extension Service, University of Florida. Available at: http://edis.ifas.ufl.edu/PD016.

Focus Groups

American Library Association. 1996. *Beyond Bean Counting, Focus Groups and Other Client-Centered Methodologies.* Audio cassette. Chicago, Ill.: ALA. Distributed by Teach'em.

Chase, Lynne C., and Jaquelina E. Alvarez. 2000. Internet Research: The Role of the Focus Group. *Library and Information Science Research* 22(4):357-369.

Connaway, Lynn Silipigni. 1996. Focus Group Interviews: A Data Collection Methodology for Decision Making. *Library Administration and Management* 10:231-239.

Connaway, Lynn Silipigni, Debra Wilcox Johnson, and Susan E. Searing. 1997. Online Catalogs From the Users' Perspective: The Use of Focus Group Interviews. *College and Research Libraries* 58(5):403-420.

Edmund, Holly, and the American Marketing Association. 1999. *The Focus Group Research Handbook.* Lincolnwood, Ill.: NTC Business Books.

Glitz, Beryl. 1998. *Focus Groups for Libraries and Librarians.* New York: Forbes.

Glitz, Beryl. 1997. The Focus Group Technique in Library Research: An Introduction. *Bulletin of the Medical Library Association* 85(4):385-390.

Greenbaum, Thomas L. 1998. *The Handbook for Focus Group Research,* second ed. Thousand Oaks, Calif.: Sage Publications.

Krueger, Richard A. 1998. *Developing Questions for Focus Groups.* Focus Group Kit Series. Thousand Oaks, Calif.: Sage Publications.

Krueger, Richard A. 1998. *Moderating Focus Groups.* Focus Group Kit Series. Thousand Oaks, Calif.: Sage Publications.

Krueger, Richard A. 1998. *Analyzing and Reporting Focus Group Results.* Focus Group Kit Series. Thousand Oaks, Calif.: Sage Publications.

Krueger, Richard A. 1994. *Focus Groups: A Practical Guide for Applied Research,* second ed. Thousand Oaks, Calif.: Sage Publications.

Krueger, Richard A., and Jean A. King. 1998. *Involving Community Members in Focus Groups.* Focus Group Kit Series. Thousand Oaks, Calif.: Sage Publications.

Marczak, Mary, and Meg Sewel. No date. Using Focus Groups for Evaluation. *CYFERNet Evaluation.* Available at: http://Ag.Arizona.Edu/fcr/fs/cyfar/focus.htm.

Morgan, David L. 1998. *The Focus Group Guidebook.* Focus Group Kit Series. Thousand Oaks, Calif.: Sage Publications.

Morgan, David L. 1998. *Planning Focus Groups.* Focus Group Kit Series. Thousand Oaks, Calif.: Sage Publications.

Morgan, David L., ed. 1993. *Successful Focus Groups: Advancing the State of the Art.* Newbury Park, Calif.: Sage Publications.

Morgan, David L. 1988. *Focus Groups as Qualitative Research.* Newbury Park, Calif.: Sage Publications.

Morrison, Heather G. 1997. Information Literacy Skills: An Exploratory Focus Group Study of Student Perceptions. *Research Strategies* 15:4-17.

Nasser, David L. 1988. Workshop: How to Run a Focus Group. *Public Relations Journal* 44(3):33-34.

Steward, D.W., and P.N. Shamdasani. 1990. *Focus Groups: Theory and Practice.* Applied Social Research Methods Series. Vol. 20. Newbury Park, Calif.: Sage Publications.

Templeton, Jane Farley. 1994. *The Focus Group: A Strategic Guide to Organizing, Conducting, and Analyzing the Focus Group Interview.* Revised edition. Chicago, Ill.: Probus Publishing Co.

Survey Questionnaires

Acquisitions Policy: Who Buys What and Why? 2000. *Library Association Record* 102(8):432.

Allen, Frank R. 1996. Materials Budgets in the Electronic Age: A Survey of Academic Libraries. *College and Research Libraries* 57(2):133.

Allen, Mary Beth. 1993. International Students in Academic Libraries: A User Survey. *College and Research Libraries* 54(4):323-324.

ASA Series: What Is a Survey? *Social Science Information Gateway.* Available at; http://www.amstat.org/sections/srms/brochures/survwhat.html.

Ashcroft, Linda, and Colin Langdon. 1999. The Case for Electronic Journals. *Library Association Record* 101(12):706-707.

Atlas, Michel C., and Melissa A. Laning. 1999. Professional Journal Reading: A Survey of Kentucky Academic Librarians. *Kentucky Libraries* 63(1):16-21.

Bancroft, Audrey F., Vicki F. Croft, and Robert Speth. 1998. A Forward-Looking Library Use Survey: WSU Libraries in the 21st Century. *The Journal of Academic Librarianship* 24(3):216-224.

Bao, Xue-Ming. 2000. Academic Library Home Pages: Link Location and Database Provision. *The Journal of Academic Librarianship* 26(3):191-195.

Benaud, Claire-Luise, Sever Michael Bordeianu, and Mary Ellen Hanson. 1999. Cataloging Production Standards in Academic Libraries. *Technical Services Quarterly* 16(3):43-67.

Brown, Linda A., and John Harper Forsyth. 1999. The Evolving Approval Plan: How Academic Librarians Evaluate Services for Vendor Selection and Performance. *Library Collections Acquisitions, and Technology Services* 23(3):231-277.

Buttlar, Lois J., and Rajinder Garcia. 1998. Catalogers in Academic Libraries: Their Evolving and Expanding Roles. *College & Research Libraries* 59(4):311-321.

Button, Leslie Horner. 2000. Impact of Bundled Databases on Serials Acquisitions in Academic Libraries. *The Serials Librarian* 38(3/4):213-218.

Calvert, Philip J. 1997. Surveying Service Quality Within University Libraries. *The Journal of Academic Librarianship* 23:408-415.

Coffta, Michael, and David M. Schoen. 2000. Academic Library Web Sites as a Source of Interlibrary Loan Lending Information: A Survey of Four- and Five-year Colleges and Universities. *Library Resources & Technical Services* 44(4):196-200.

Doyle, Christine. 1995. The Perceptions of Library Service Questionnaire PLSQ: The Development of a Reliable Instrument to Measure Student Perceptions of and Satisfaction with Quality of Service in an Academic Library. *New Review of Academic Librarianship* 1:139-154.

East, John W. 2001. Academic Libraries and the Provision of Support for Users of Personal Bibliographic Software: A Survey of Australian Experience with Endnote. *LASIE* 32(1):64-70.

Evans, Geraint, and Jane Del-Pizzo. 1999. "Look, Hear, Upon This Picture": A Survey of Academic Users of the Sound and Moving Image Collection of the National Library of Wales. *Journal of Librarianship and Information Science* 31(3):152-167.

Hart, Richard L. 2000. Co-authorship in the Academic Library Literature: A Survey of Attitudes and Behaviors. *The Journal of Academic Librarianship* 26(5):339-345.

Hernon, Peter, and Philip J. Calvert. 1996. Methods for Measuring Service Quality in University Libraries in New Zealand. *The Journal of Academic Librarianship* 22:387-391.

Herring, Susan Davis. 2001. Using the World Wide Web for Research: Are Faculty Satisfied? *The Journal of Academic Librarianship* 27(3):213-219.

Hoffman, Irene M., Amy Sherman Smith, and Leslie DeBonae. 2000. Factors for Success: Academic Library Development Survey Results. *Library Trends* 48(3):540-559.

Ikeuchi, Atsushi. 1998. Dimensions of Academic Library Effectiveness. *Library and Information Science* 39:1-29.

Janes, Joseph, David S. Carter, and Patricia Memmott. 1999. Digital Reference Services in Academic Libraries. *Reference & User Services Quarterly* 39(2):145-150.

Johnson, Denise J. 1997. Merging? Converging? A Survey of Research and Report on Academic Library Reorganization and the Recent Rash of Marriages Between Academic Libraries and University Computer Centers. *Illinois Libraries* 79(2):61.

Julien, Heidi E. 2000. Information Literacy Instruction in Canadian Academic Libraries: Longitudinal Trends and International Comparisons. *College & Research Libraries* 61(6):510-523.

Kwak, Gail Stern. 2000. Government Information on the Internet: Perspectives from Louisiana's Academic Depository Libraries. *Louisiana Libraries* 63(2):17-22.

Libby, Katherine A. 1997. A Survey on the Outsourcing of Cataloging in Academic Libraries. *College and Research Libraries* 58(6):550.

Love, Christine, and John Feather. 1998. Special Collections on the World Wide Web: A Survey and Evaluation. *Journal of Librarianship and Information Science* 30(4):215-222.

Maughan, Patricia Davitt. 1999. Library Resources and Services: A Cross-Disciplinary Survey of Faculty and Graduate Student Use and Satisfaction. *Journal of Academic Librarianship* 25(5):354.

Norman, O. Gene. 1997. The Impact of Electronic Information Sources on Collection Development: A Survey of Current Practice. *Library Hi Tech* 15(1-2):123-132.

Organ, M., and M. Janatti. 1997. Academic Library Seating: A Survey of Usage, with Implications for Space Utilization. *Australian Academic and Research Libraries, AARL* 28(3):205.

Pavy, Jeanne A. 2000. Special Collections in Regional Academic Libraries: An Informal Survey. *Louisiana Libraries* 63(2):14-16.

Perkins, Gay Helen, and Haiwang Yuan. 2000. Genesis of a Web-based Satisfaction Survey in an Academic Library: The Western Kentucky University Libraries' Experience. *Library Administration & Management* 14(3):159-166.

Porter, Gayle, and Paul Bredderman. 1997. Nonprint Formats: A Survey of the Work and Its Challenges for the Cataloger in ARL Academic Libraries. *Cataloging and Classification Quarterly* 24(3-4):125.

Rasinski, Kenneth A., David Mingay, and Norman M. Bradburn. 1994. Do Respondents Really "Mark All That Apply" on Self-administered Questions? *Public Opinion Quarterly* 58(3):400-408.

Ren, Wen-Hua. 2000. Library Instruction and College Student Self-efficacy in Electronic Information Searching. *The Journal of Academic Librarianship* 26(5):323-328.

Rich, Linda A., and Julie L. Rabine. 1999. How Libraries Are Providing Access to Electronic Serials: A Survey of Academic Library Web Sites. *Serials Review* 25(2):35-46.

Sanchez, Maria Elena. 1992. Effects of Questionnaire Design on the Quality of Survey Data. *Public Opinion Quarterly* 56(2):206-217.

Schilling, Katherine, and Charles B. Wessel. 1995. Reference Librarians' Perceptions and Use of Internet Resources: Results of a Survey of Academic Health Science Libraries. *Bulletin of the Medical Library Association* 83(4):509.

Schneider, Tina M. 2001. The Regional Campus Library and Service to the Public. *The Journal of Academic Librarianship* 27(2):122-127.

Scigliano, Marisa. 2000. Serial Use in a Small Academic Library: Determining Cost-Effectiveness. *Serials Review* 26(1):43.

Shemberg, Marian, and Cheryl R. Sturko Grossman. 1999. Electronic Journals in Academic Libraries: A Comparison of ARL and Non-ARL Libraries. *Library Hi Tech* 17(1):26-45.

Shonrock, Diana D., ed. 1996. *Evaluating Library Instruction: Sample Questions, Forms, and Strategies for Practical Use*. Chicago, Ill.: American Library Association.

Sinn, Robin N. 1999. A Comparison of Library Instruction Content by Biology Faculty and Librarians. *Research Strategies* 17(1):23-34.

Talbot, Dawn E., Gerald R. Lowell, and Kerry Martin. 1998. From the User's Perspective—The UCSD Libraries User Survey Project. *Journal of Academic Librarianship* 24(5):357.

Travica, B. 1997. Organizational Aspects of the Virtual/Digital Library: A Survey of Academic Libraries. *Proceedings of the ASIS Annual Meeting* 34:149–161. Medford, N.J.: Information Today.

Travica, Bob. 1999. Organizational Aspects of the Virtual Library: A Survey of Academic Libraries. *Library & Information Science Research* 21(2):173-203.

White, Marilyn Domas. 2001. Diffusion of an Innovation: Digital Reference Service in Carnegie Foundation Master's Comprehensive Academic Institution Libraries. *The Journal of Academic Librarianship* 27(3):173-187.

Wu, Huey-meei. 1998. Academic Library User Survey: National Chiao-Tung University Case Study. *Journal of Educational Media & Library Sciences* 36(2):171-196.

Transaction Log Analysis

Atlas, Michael C., Karen R. Little, and Michael O. Purcell. 1997. Flip Charts at the OPAC: Using Transaction Log Analysis to Judge Their Effectiveness at Six Libraries of the University of Louisville. *Reference and User Services Quarterly* 37(1):63-69.

Bangalore, Nirmala S. 1997. Re-Engineering the OPAC Using Transaction Logs at the University of Illinois at Chicago. *Libri* 47:67-76.

Blecic, Deborah D., Nirmala S. Bangalore, and Josephine L. Dorsch. 1998. Using Transaction Log Analysis to Improve OPAC Retrieval Results. *College and Research Libraries* 59:39-50.

Ciliberti, Anne C., Marie L. Radford, and Gary P. Radford. 1998. Empty Handed? A Material Availability Study and Transaction Log Analysis Verification. *The Journal of Academic Librarianship* 24(4):282-289.

Connaway, Lynn Silipigni, John Budd, and Thomas R. Kochtanek. 1995. An Investigation of the Use of an Online Catalog: User Characteristics and Transaction Log Analysis. *Library Resources and Technical Services* 39:142-152.

Ferl, Terry Ellen, and Larry Millsap. 1996. The Knuckle-Cracker's Dilemma: A Transaction Log Study of OPAC Subject Searching. *Information Technology and Libraries* 15:81-98.

Jones, Susan, Mike Gatford, and Thien Do. 1997. Transaction Logging. *Journal of Documentation* 53:35-50.

Kaske, Neal K. 1993. Research Methodologies and Transaction Log Analysis: Issues, Questions, and a Proposed Model. *Library Hi Tech* 11(2):79-86.

King, Natalie Schoch. 1993. End-User Errors: A Content Analysis of PaperChase Transaction Logs. *Bulletin of the Medical Library Association* 81:439-441.

Kurth, Martin. 1993. The Limits and Limitations of Transaction Log Analysis. *Library Hi Tech* 11(2):98-104.

Library Hi Tech. 1993. Special Issue on Transaction Log Analysis 11(2).

Millsap, Larry, and Terry Ellen Ferl. 1993. Search Patterns of Remote Users: An Analysis of OPAC Transaction Logs. *Information Technology and Libraries* 12:321-343.

Peters, Thomas A. 1996. Using Transaction Log Analysis for Library Management Information. *Library Administration and Management* 10:20-25.

Peters, Thomas A. 1993. The History and Development of Transaction Log Analysis. *Library Hi Tech* 11(2):41-66.

Peters, Thomas A., Martin Kurth, Patricia Flaherty, et al. 1993. An Introduction to the Special Section on Transaction Log Analysis. *Library Hi Tech* 11(2):38-40.

Sandore, Beth. 1993. Applying the Results of Transaction Log Analysis. *Library Hi Tech* 11(2):87-97.

Sandore, Beth, Patricia Flaherty, Neal K. Kaske, et al. 1993. A Manifesto Regarding the Future of Transaction Log Analysis. *Library Hi Tech* 11(2):105-106.

Shelfer, Katherine M. 1998. *Transaction Log Analysis as a Method to Measure the Frequencies of Searcher Access to Library Research Guides*. Madison, Wisc.: University of Wisconsin Press.

Sullenger, Paula. 1997. A Serials Transaction Log Analysis. *Serials Review* 23(3):21-26.

Wallace, Patricia M. 1993. How Do Patrons Search the Online Catalog When No One's Looking? Transaction Log Analysis and Implications for Bibliographic Instruction and System Design. *RQ* 33:239-252.

Wyly, Brendan J. 1996. From Access Points to Materials: A Transaction Log Analysis of Access Point Value for Online Catalog Users. *Library Resources and Technical Services* 40:211-236.

Usability Testing Including Protocol Analysis and Heuristic Evaluation

Allen, Bryce L. 1994. Cognitive Abilities and Information System Usability. *Information Processing and Management* 30:177-191.

Anderson, Theresa. 1999. Searching for Information: Applying Usability Testing Methods to a Study of Information Retrieval and Relevance Assessment. *Australian Academic & Research* Libraries 30(3):189-199.

Baker, Gayle S., and Flora G. Shrode. 1999. A Heuristic Approach to Selecting Delivery Mechanisms for Electronic Resources in Academic Libraries. *Journal of Library Administration* 26(3-4):153-167.

Battleson, Brenda, Austin Booth, and Jane Weintrop. 2001. Usability Testing of an Academic Library Web Site: A Case Study. *The Journal of Academic Librarianship* 27(3):188-198.

Branch, Jennifer. 2000. Investigating the Information-Seeking Processes of Adolescents: The Value of Using Think Alouds and Think Afters. *Library and Information Science Research* 22(4):371-392.

Buttenfield, Barbara Pfeil. 1999. Usability Evaluation of Digital Libraries. *Science and Technology Libraries* 17(3-4):39-59.

Campbell, Nicole. 1999. Discovering the User: A Practical Glance at Usability Testing. *The Electronic Library* 17(5):307-311.

Chance, Toby. 1993. Ensuring Online Information Usability. *The Electronic Library* 11:237-239.

Chisman, Janet K., Karen R. Diller, and Sharon L. Walbridge. 1999. Usability Testing: A Case Study. *College and Research Libraries* 60(6):456-461.

Crerar, A., and D. Benyon. 1998. Integrating Usability into Systems Development. In *The Politics of Usability: A Practical Guide to Designing Usable Systems in Industry*, edited by Lesley Trenner and Joanna Bawa. London: Springer-Verlag.

Dickstein, Ruth, and Victoria A. Mills. 2000. Usability Testing at the University of Arizona Library: How to Let the Users In On the Design. *Information Technology and Libraries* 19(3):144-151.

Dillon, Andrew. 2001. Artifacts as Theories: Convergence Through User-Centered Design. *Selected Online Publications by Andrew Dillon.* Indiana University. Available at: http://www.slis.indiana.edu/adillon/web/rescont.html.

Duncker, Elke, Yin Leng Theng, and Norlisa Mohd-Nasir. 2000. Cultural Usability in Digital Libraries. *Bulletin of the American Society for Information Science* 26(4):21-22.

Dykstra, D.J. 1993. *A Comparison of Heuristic Evaluation and Usability Testing: The Efficacy of a Domain-Specific Heuristic Checklist.* Ph.D. dissertation. Department of Industrial Engineering, Texas A&M University, College Station, Texas.

Fichter, Darlene. 2000. Head Start: Usability Testing Up Front. *Online* 24(1):79-81.

Fichter, Darlene. 2001. Testing the Web Site Usability Waters. *Online* 25(2):78-80.

Gluck, Myke. 1998. The Application of the Usability Approach in Libraries and Information Centers for Resource Selection and Deployment. *Journal of Education for Library and Information Science* 39(2):90-99.

Gullikson, Shelley, Ruth Blades, and Marc Bragdon. 1999. The Impact of Information Architecture on Academic Web Site Usability. *The Electronic Library* 17(5):293-304.

Head, Alison J. 1999. Web Redemption and the Promise of Usability. *Online* 23(6):20-23.

Head, Alison J. 1997. Web Usability and Essential Interface Design Issues. *Proceedings of the 18th National Online Meeting,* New York, May 13-15, 1997.

Hert, Carol Ann, Elin K. Jacob, and Patric Dawson. 2000. A Usability Assessment of Online Indexing Structures in the Networked Environment. *Journal of the American Society for Information Science* 51(1):971-988.

Hudson, Laura. 2000. Radical Usability or, Why You Need to Stop Redesigning Your Web Site. *Library Computing* 19(1/2):86-92.

Jeffries, R., et al. 1991. User Interface Evaluation in the Real World: A Comparison of Four Techniques. In *Proceedings ACM CHI '91 Conference,* edited by S. P. Robertson, G. M. Olson, and J. S. Olson. New York: Association for Computing Machinery.

Levi, M.D., and F.G. Conrad. 2000. Usability Testing of World Wide Web Sites. Bureau of Labor Statistics Research Papers. Available at: http://stats.bls.gov/ore/htm_papers/st960150.htm.

McGillis, Louise, and Elaine G. Toms. 2001. Usability of the Academic Library Web Site: Implications for Design. *College and Research Libraries* 62(4):355-368.

McLean, Stuart, Michael B. Spring, and Edie M. Rasmussen. 1995. Online Image Databases: Usability and Performance. *The Electronic Library* 13:27-42.

Molich, R., and J. Nielsen. 1990. Improving a Human-Computer Dialogue. *Communications of the ACM* 33(3):338-348.

Morgan, Eric Lease. 1999. Marketing Through Usability. *Computers in Libraries* 19(80):52-53.

Morrison, Heather G. 1999. Online Catalogue Research and the Verbal Protocol Method. Concordia University Students Conduct Research on DRA's Infogate. *Library Hi Tech* 17(2):197-206.

Nielsen, Jakob. 2001. Are Users Stupid? *Alertbox* (Feb. 4). Available at: http://www.useit.com/alertbox/20010204.html.

Nielsen, Jakob. 1998. Cost of User Testing a Website. *Alertbox* (May 3). Available at: http://www.useit.com/alertbox/980503.html.

Nielsen, Jakob. 2000. *Designing Web Usability*. Indianapolis, Ind.: New Riders Press.

Nielsen, Jakob. 1994. Enhancing the Explanatory Power of Usability Heuristics. In *CHI '94 Conference Proceedings*, edited by B. Adelson, S. Dumais, and J. Olson. New York: Association of Computing Machinery.

Nielsen, Jakob. 1992. Finding Usability Problems Through Heuristic Evaluation. In *Proceedings ACM CHI'92 Conference*, edited by P. Bauersfeld, J. Bennett, and G. Lynch. New York: Association for Computing Machinery.

Nielsen, Jakob. 2001. First Rule of Usability? Don't Listen to Users. *Alertbox* (Aug. 5). Available at: http://www.useit.com/alertbox/20010805.html.

Nielsen, Jakob. [No date.] Heuristic Evaluation. Available at: http://www.useit.com/papers/heuristic/.

Nielsen, Jakob. 1990. Paper versus Computer Implementations as Mockup Scenarios for Heuristic Evaluation. *Proceedings of IFIP IINTERACT 90, Third International Conference on Human-Computer Interaction*, edited by D. Diaper et al. Amsterdam: North-Holland.

Nielsen, Jakob. 2001. Success Rate: The Simplest Usability Metric. *Alertbox* (Feb. 18). Available at: http://www.useit.com/alertbox/20010218.html.

Nielsen, Jakob. [No date.] Ten Usability Heuristics. Available at: http://www.useit.com/papers/heuristic/heuristic_list.html.

Nielsen, Jakob. 1993. *Usability Engineering*. Boston, Mass.: Academic Press.

Nielsen, Jakob. 2001. Usability Metrics. *Alertbox*. (Jan. 21). Available at: http://www.useit.com/alertbox/20010121.html.

Nielsen, Jakob. 2000. Why You Only Need to Test With 5 Users. *Alertbox* (March 19). Available at: http://www.useit.com/alertbox/20000319.html.

Nielsen, Jakob, T.K. Landauer. 1993. A Mathematical Model of the Finding of Usability Problems. *Proceedings ACM/IFIP INTERCHI'93 Conference*, edited by S. Ashlund et al. New York: Association for Computing Machinery.

Nielsen, Jakob, Robert L. Mack, and Arthur G. Elser. 1995. Usability Inspection Methods. *Technical Communication* 42(4):661.

Nielsen, Jakob, and Robert L.Mack, eds. 1994. *Usability Inspection Methods*. New York: John Wiley and Sons.

Nielsen, Jakob, and R. Molich. 1990. Heuristic Evaluation of User Interfaces. In *Proceedings ACM CHI'90 Conference*, edited by J. Carrasco Chew, and J. Whiteside. New York: Association for Computing Machinery.

Olason, Susan C. 2000. Let's Get Usable! Usability Studies for Indexes. *The Indexer* 22(2):91-95.

Pack, Thomas. 2001. Use It or Lose It: Jakob Nielsen Champions Content Usability. *Econtent* 24(4):44-46.

Palmquist, Ruth Ann. 2001 An Overview of Usability for the Study of User's Web-Based Information Retrieval Behavior. *Journal of Education for Library and Information Science* 42(2):123-136.

Park, Soyeon. 2000. Usability, User Preferences, Effectiveness, and User Behaviors When Searching Individual and Integrated Full-Text Databases: Implications for Digital Libraries. *Journal of the American Society for Information Science* 51(5):456-468.

Prasse, M. J. and R. Tigner. 1992. The OCLC Usability Lab: Description and Methodology. In *13th National Online Meeting Proceedings—1992, New York, May 5-7, 1992*, edited by Martha E. Williams. Medford, N.J.: Learned Information, Inc.

Rousseau, G.K. 1999. Assessing the Usability of On-line Library Systems. *Communication Abstracts* 22(2):536.

Rubin, Jeffrey. 1994. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. New York: John Wiley and Sons.

Sakaguchi, Kazuko. 1999. Journal Assessment: A Protocol Analysis for Establishing Criteria for Shared Responsibilities for the Japanese Vernacular Journal Collection. *Journal of East Asian Libraries*, 118:1-20.

Veldof, Jerilyn R., Michael J. Prasse, and Victoria A. Mills. 1999. Chauffeured by the User: Usability in the Electronic Library. *Journal of Library Administration* 26(3-4):115-140.

Walbridge, Sharon L. 2000. Usability Testing and Libraries: The WSU Experience. [Washington State University]. *Alki* 16(3):23-24.

Wiedenbeck, Susan, Robin Lampert, and Jean Scholtz. 1989. Using Protocol Analysis to Study the User Interface. *Bulletin of the American Society for Information Science* 15:25-26.

Zhang, Zhijun, Victor Basili, and Ben Schneiderman. 1999. Perspective-Based Usability Inspection: An Empirical Validation of Efficacy. *Empirical Software Engineering* 4(1):43-69.

APPENDIX B
Participating Institutions

California Digital Library
Carnegie Mellon University
Columbia University
Cornell University
Emory University
Harvard University
Indiana University
Johns Hopkins University
Library of Congress
New York Public Library
North Carolina State University
Pennsylvania State University
Stanford University
University of Chicago
University of Illinois
University of Michigan
University of Minnesota
University of Pennsylvania
University of Southern California
University of Texas
University of Tennessee
University of Virginia
University of Washington
Yale University

APPENDIX C
Survey Questions

1. What data do you gather to assess user needs and the use and usability of your library?
2. How do you gather the data?
3. How do you analyze the data?
4. Why are you gathering and analyzing the data?
5. How do you use the results of the data analysis?
6. How does the process seem to work? What works well? What doesn't work so well?
7. How would you change the process?

APPENDIX D
Traditional Input, Output, and Outcome Measures

The body of this report focuses on studies of users and electronic resource usage because these were the areas that the Digital Library Federation (DLF) survey respondents spent most of their time discussing during the interviews. Putting these issues in the foreground, however, is somewhat misleading, because libraries have traditionally gathered and continue to gather statistics related to the size, use, and impact of all of their collections and services. These traditional measures are being expanded to embrace digital library activities in order to capture the full scope of library performance. This expansion is problematic for reasons already acknowledged; for example, because libraries are in transition and standard definitions and reporting mechanisms are not yet fully established. Nevertheless, substantial progress is being made through the efforts of groups such as the Association of Research Libraries (ARL), which are undertaking large projects to field-test and refine new measures.

This appendix describes what DLF respondents reported about their input, output, and outcome measures to indicate the full scope of their assessment practices and to provide a context in which to interpret both the design and the results of the user and usage studies presented in the body of this report. The treatment is uneven in detail because the responses were uneven. Many respondents talked at great length about some topics, such as the use of reference services. In other cases, respondents mentioned a measure and brushed over it in a sentence. The unevenness of the discussion suggests where major difficulties or significant activity exists. As much as possible, the approach follows that used in the body of this report: What is the measure? Why is it gathered? How are the data used? What challenges do libraries face with it?

1. INPUT AND OUTPUT MEASURES

Traditional measures quantify a library's raw materials or potential to meet user needs (*inputs*) and the actual use of library collections and services (*outputs*). Input and output statistics reveal changes in what libraries do over time. For example, they provide a longitudinal look at the number of books purchased and circulated per year. Traditional approaches to measuring inputs and outputs focus on physical library resources. Libraries are slowly building a consensus on what to measure and how to measure inputs and outputs in the digital environment. The goal is standard definitions that facilitate gathering digital library data that can be compared with traditional library data from their own institution and from others. Developing such standards is difficult for many reasons, not the least of which is the basic fact of digital library life addressed in the transaction log analysis section of this report: much of the data are provided by vendor systems or software packages that capture and count transactions differently and do not always provide the statistics that libraries prefer. Though the form of the problem is new in the sense that the data are provided by units not controlled by the library, the problem itself is not. Even in the traditional library environment, definitions were not uniform. Comparison and interpretation were complicated by contextual factors such as the length of circulation loan periods and institutional missions that shaped library statistics and performance.

1.1. Input Measures: Collection, Staff, and Budget Sizes

Libraries have traditionally gathered statistics and monitored trends in the size of their collections, staff, and budgets. Collection data are gathered in an excruciating level of detail; for example, the number of monographs, current serials, videos and films, microforms, CDs, software, maps, musical scores, and even the number of linear feet of archival materials. The data are used to track the total size of collections and collection growth per year. Typically, the integrated library management system (ILS) generates reports that provide collection data. Staff sizes are traditionally tracked in two categories: professionals (librarians) and support staff. The library's business manager or human resources officer provides these data. The business manager tracks budgets for salaries, materials, and general operation of the library. DLF respondents indicated that collection, staff, and budget data are used primarily to meet reporting obligations to national organizations such as ARL and ACRL, which monitor library trends. Ratios are compiled to assess such things as the number of new volumes added per student or full-time faculty member, which reveals the impact of the economic crisis in scholarly communication on library collections.

New measures are being developed to capture the size of the digital library as an indication of the library's potential to meet user

needs for electronic resources. DLF respondents reported using the following digital library input measures:

- Number of links on the library Web site
- Number of pages in the library Web site
- Number of licensed and locally maintained databases
- Number of licensed and locally maintained e-journals
- Number of licensed and locally maintained e-books
- Number of locally maintained digital collections
- Number of images in locally maintained digital collections
- Total file size of locally maintained databases and digital collections

Whether libraries also count the number of e-journals, e-books, or digital collections that they link to for free is unclear. Some of these measures can be combined with traditional collection statistics to reveal the libraries' total collection size (for example, the number of physical monographs plus the number of e-books) and trends in electronic collection growth. DLF respondents indicated that they were beginning to capture the following composite performance measures:

- Percentage of book collection available electronically
- Percentage of journal collection available electronically
- Percentage of reserves collection available electronically
- Percentage of the materials budget spent on e-resources

In many cases, baseline data are being gathered. Little historical data are available to assess trends within an institution. Even if multiyear data are available, libraries have had no way to compare their efforts with those of their peer institutions, because there is no central reporting mechanism for digital library input measures. ARL will soon begin gathering such e-metrics, but other reporting organizations appear to be further behind in this regard.

DLF respondents talked about the difficulty of compiling these data. The data reside in different units within the library, and the systems that these units use do not support this kind of data gathering and reporting. The upshot is a labor-intensive effort to collect, consolidate, and manage the statistics. ARL's E-Metrics Phase II Report, *Measures and Statistics for Research Library Networked Services*, describes the related issue of "the organizational structure needed to manage electronic resources and services, particularly the configuration of personnel and workflow to support the collection of statistics and measures."[1] Interpreting these data is also an issue. For example, what does it mean if the number of pages on the library Web site shrinks following a major redesign of the site? Just as traditional input measures seemed to assume that more books were better than fewer books, should libraries assume that more Web pages are necessarily better than fewer Web pages? DLF respondents didn't think so.

1 http://www.arl.org/stats/newmeas/emetrics/phasetwo.pdf, October 2001, p. 41.

User studies and an interpretive framework based on a study of key factors in the larger environment are needed to interpret the data.

Some DLF respondents commented on trends in staff and budget sizes. They talked about hiring more technical staff (technicians, system managers, programmers) and other personnel (interface designers, human factors researchers) needed to support digital library initiatives. These positions are funded primarily by eliminating open positions because personnel budgets do not accommodate adding positions. At the time the DLF interviews were conducted, there was a crisis in hiring information technology (IT) personnel in higher education because salaries were not competitive with those in the corporate sector.[2] The situation was even more urgent for academic libraries, which often could not compete with IT salaries even within their institution. The recent folding of many dot-coms might make higher education salaries more competitive and facilitate filling these positions, but unless the inequity in IT salaries within an institution is addressed, libraries could continue to have problems in this area. DLF respondents commented that materials budgets did not keep pace with the rising cost of scholarly communications, and that operating or capital budgets were often inadequate to fund systematic replacement cycles for equipment, not to mention the purchase of new technologies.

1.2. Output Measures

Libraries have traditionally gathered statistics and monitored trends in the use of their collections and services. They often compare traditional usage measurements across institutions, although these comparisons are problematic because libraries, like vendors, count different things and count the same things in different ways. Though settling for "good-enough" data seems to be the mantra of new measures initiatives and conferences on creating a "culture of assessment," libraries have apparently been settling for good-enough data since the inception of their data gathering. Reference service data are a case in point, described in section 1.2.4. of this appendix. The following discussion of output measures reflects the expansion of traditional measures to capture the impact of digital initiatives on library use and the issues and concerns entailed in this expansion.

1.2.1. Gate Counts

Gate counts indicate the number of people who visit the physical library. Students often use an academic library as a place for quiet study, group study, or even social gatherings. Capturing gate counts is a way to quantify use of the library building apart from use of library collections and services. Libraries employ a variety of technological devices to gather gate counts. The data are often gathered at the point of exit from the library and compiled at different time peri-

[2] Recruiting and Retaining Information Technology Staff in Higher Education. Available at: http://www.educause.edu/pub/eb/eb1.html. August 2000.

ods throughout the day. Depending on the device capabilities, staff might manually record gate count data on a paper form at specified times of the day and later enter it into a spreadsheet to track trends.

Libraries include gate count data in annual reports. They use gate counts to adjust staffing and operating hours, particularly around holidays and during semester breaks. Sites capturing the data with card-swipe devices can use the data to track usage patterns of different user communities.[3] One DLF respondent reported that regression analysis of exit data can explain fluctuations in reference activity and in-house use of library materials. If one of these variables is known, the other two can be statistically estimated. However, no library participating in the DLF survey reported using gate counts to predict reference service or in-house use of library materials. Adjustments to staffing and operating hours appear to be made based on gross gate counts at different time periods of the day and on the academic and holiday calendar. Gate count data, like data from many user studies, appear to be gathered in some cases even though libraries do not have the will, organizational capacity, skill, or interest to mine, interpret, and use them effectively in strategic planning.

Digital library initiatives introduce a new dimension to visiting the library. The notion of a "virtual" visit raises issues of definition, guidelines for how to gather the data, and how or whether to compile traditional gate counts and virtual visits as a composite measure of library use. Is a virtual visit a measure of use of the library Web site, the OPAC, or an electronic resource or service? All of the above? Surely it is not a matter of counting every transaction or page fetched, in which case a definition is needed for what constitutes a "session" in a stateless, sessionless environment such as unauthenticated use of Web resources. The recommendation in the ARL E-Metrics Phase II Report and the default in some Web transaction analysis software define a session based on a 30-minute gap of inactivity between transactions from a particular IP address.[4] Compiling a composite measure of traditional gate counts and virtual visits introduces a further complication, because virtual visits from IP addresses within the library must be removed from the total count of virtual visits to avoid double counting patrons who enter the physical library and use library computers to access digital resources.

Libraries are struggling with how to adjudicate these issues and determine what their practice will be. Their decisions are constrained by what data it is possible and cost-effective to gather. One DLF site has decided to define virtual visits based strictly on use of the library Web site, a 30-minute gap of inactivity from an IP address, and ag-

[3] Card-swipe exit data capture user IDs, which library the user is in, and the date and time. IDs can be mapped to demographic data in the library patron database to determine the users' status and school (e.g., graduate student, business school).

[4] http://www.arl.org/stats/newmeas/emetrics/phasetwo.pdf. October 2001, pp. 66-67.

gregate data on virtual visits inside and outside of the libraries. Given their equipment replacement cycle and the number of new machines and hence new IP addresses deployed each year in the library, this library decided that the benefits of calculating the number of virtual visits from machines inside the library did not warrant the costs.

1.2.2. Circulation and In-House Use

Circulation statistics traditionally indicate how many items were checked out to users or used within the library. Circulation data reports are generated routinely from the Integrated Library System (ILS). Initial checkouts and renewals are tracked separately because national surveys require it. Reshelving data, gathered manually or through the ILS, are used to assess in-house use of library materials. Items that circulate through other venues, for example, analog or digital slides, might not be included in circulation statistics.

Libraries include circulation data in annual reports and national library surveys. The data are used to:

- Identify items that have never circulated and inform retention and cancellation decisions
- Assess or predict book use to help decide what to move to off-site storage[5]
- Decide whether the appropriate materials are in off-site storage
- Determine staffing at the circulation desk by examining patterns of circulation activity per hour, day, and academic quarter

In addition, one DLF respondent mentioned conducting a demographic analysis of circulation data to determine circulation per school, user status, library, and subject classification. The results were used to inform collection development decisions. Other DLF respondents simply commented that they know that humanists use books and scientists use journals.

Libraries also generate financial reports of fines and replacement costs for overdue and lost books. The data are tracked as a source of important revenue and are frequently used to help fund underbudgeted student employee wages. Collection developers determine whether lost books will be replaced, presumably based on a cost-benefit analysis of the book's circulation and replacement cost. Some DLF respondents also reported tracking recalls and holds, but did not explain how these data are used. If the data are used to track user demand for particular items and inform decisions about whether to purchase additional copies, they serve a purpose. If the data are not used, data collection is purposeless.

The digital environment also introduces a new dimension to circulation data gathering, analysis, and use. For example, a comprehensive picture of library resource use requires compiling data on use of traditional (physical) and digital monographs and journals.

[5] For example, see Craig Silverstein and Stuart M. Shieber. 1996. Predicting Individual Book Use for Off-Site Storage Using Decision Trees. *Library Quarterly* 66(3):266-293.

Usage data on electronic books and journals are not easily gathered and compiled because they are not checked out or re-shelved in the traditional sense and because the data are for the most part provided by vendors—in different formats and time periods, and based on different definitions. Ideally, use of all physical and digital resources would be compiled, including use of physical and digital archival materials, maps, and audio and video resources. The discussions of transaction log analysis and virtual visits earlier in this report describe many of the difficulties inherent in tracking "circulation" or "in-house use" of electronic resources. A few DLF respondents mentioned efforts to compile book and journal data as their foray into this area, but a comprehensive picture of use of library collections appears to be a long way off.

1.2.3. Reserves

Faculty put items that they want students to use, but do not distribute in class or require them to purchase, on reserve in the library. Libraries track reserve materials in great detail. Reserves are tracked as both input and output measures. Both dimensions are treated here to facilitate an understanding of the complexity of the issues. Libraries place items on course reserves in traditional paper and electronic formats. Some DLF sites operate dual systems, offering both print and e-reserves for the same items. DLF respondents reported tracking the following:

- The number of items on reserve in traditional and digital format
- The use of traditional and e-reserve items
- The percentage of reserve items available electronically
- The percentage of reserve use that is electronic

The number of traditional and digital reserve items in some cases is tracked manually because the ILS cannot generate the data. Depending on how reserves are implemented, use of traditional reserves (for example, books and photocopies) might be tracked by the circulation system. Tracking use of e-reserves requires analysis of Web server logs (for example, the number of PDF files downloaded or pages viewed). The data are used to track trends over time, including changes in the percentage of total reserve items available electronically and the percentage of total reserve use that is electronic. Data on reserve use may be included in annual reports.

One DLF site reported analyzing Web logs to prepare daily and hourly summaries of e-reserves use, including what documents users viewed, the number of visits to the e-reserves Web site, how users navigated to the e-reserves Web site (from what referring page), and what Web browser they used. This library did not explain how these data are used. Another site reported tracking the number of reserve items per format using the following format categories: book, photocopy, personal copy, and e-reserves. Their e-reserve collection does not include books, so to avoid comparing apples with oranges, they calculate their composite performance measures without including books in the count of traditional reserve items or use. Several sites

provide or plan to provide audio or video e-reserves. Only time will tell if they begin to track formats within e-reserves and how this will affect data gathering and analysis.

DLF respondents also mentioned tracking the following information manually:

- The number of reserve items per academic department, faculty member, and course number
- The number of requests received per day to put items on reserve
- The number of items per request
- The number of items made available on reserves per day
- The number of work days between when the request was submitted and when the items are made available on reserves
- The number of pages in e-reserve items

Data about the number of requests per day, the number of items per request, and the amount of time that passes between when a request is placed and when the item becomes available on reserve are used to estimate workload, plan staffing, and assess service quality. The number of pages in e-reserve items is a measure of scanning activity or digital collection development. It is also used as the basis for calculating e-resource use in systems where e-reserves are delivered page by page. (The total number of e-reserve page hits is divided by the average number of pages per e-reserve item to arrive at a measure comparable to checkout of a traditional reserve item.) No indication was given for how the data on reserve items per department, faculty, and course were used. If converted to percentages, for example, the percentage of faculty or departments requesting reserves, the data would provide an indication of market penetration. If, however, the data are not used, data collection is purposeless.

1.2.4. Reference

Reference data are difficult to collect because reference service is difficult to define, evolving rapidly, and being offered in new and different ways. The problem is compounded because naturally the methods for assessing new service delivery evolve at a slower rate than the service forms themselves do. DLF respondents reported offering reference service in the following ways, many of which are online attempts to reach remote users:

- Face-to-face at the reference desk
- Telephone at the reference desk
- Telephone to librarian offices
- E-mail, using a service e-mail address or Web-based form on the library's Web site
- E-mail directly to librarians
- U.S. Postal Service
- Chat software
- Virtual Reference Desk software
- Teleconferencing software

Libraries are also collaborating to provide online or digital reference service. For example, some DLF sites are participating in the Collaborative Digital Reference Service,[6] which is a library-to-library service to researchers available any time, anywhere, through a Web-based, international network of libraries and other institutions organized by the Library of Congress. Other collaborative digital reference services include the 24/7 Reference Project and the Virtual Reference Desk Network.[7] The DLF, OCLC, and other organizations are supporting a study of online reference services being conducted by Charles McClure and David Lankes. Findings from the study so far reveal a wide range of concerns and need for new measures. For example, there are concerns about competitive reference services in the commercial sector, concerns about decreasing traditional reference statistics and the potential volume of digital reference questions, and a need for instruments to measure the effectiveness, efficiency, costs, and outcomes of digital reference.[8]

Most DLF libraries track reference data, but they define different categories of questions to count, and they count at different frequencies. At bare minimum, libraries count questions asked at the reference desk and distinguish "reference" questions from "directional" questions. Some libraries distinguish "quick reference" questions from "real reference" questions. Some libraries explicitly count and categorize "technical" questions about computers, printers, or the network. Some include technical questions under the rubric of "reference" questions. Some do not count technical questions at all. Some have a category for "referrals" to other subject specialists. Some have an "Other" category that is undefined. Some libraries track the time of day and day of week questions are asked at the reference desk. Some track the school and status of the user and the reference desk location. Some libraries gather reference desk data routinely. Others sample, for example, two randomly selected days per month, two weeks per year, or two weeks per quarter. Some libraries include in their reference statistics questions that go directly to the librarian's desk via telephone or personal e-mail. Others make no effort to gather such data. Two apparently new initiatives are to track the length of reference transactions and the number of reference questions that are answered using electronic resources.

Compiling data from different venues of reference service is time-consuming because the data gathering is dispersed. Reference desk questions are tracked manually at each desk. Librarians manually track telephone and e-mail questions that come directly to them. Such manual tracking is prone to human error. E-mail questions to a reference service e-mail address are tracked on an electronic bulletin board or mailbox. Chat reference questions are counted through

[6] See http://www.loc.gov/rr/digiref.

[7] See http://www.vrd.org/network.shtml.

[8] Project updates are available at http://quartz.syr.edu/quality.

transaction log analysis. Often efforts to assemble these data are not well organized.

Despite these difficulties and anomalies, reference data are included in annual reports and national library surveys. The data are used to determine

- Performance trends over time, including the percentage of reference questions submitted electronically and the percentage of reference questions answered using electronic resources
- Appropriate hours of reference service
- Appropriate staffing at the reference desk during specific hours of the day
- Instruction to be provided for different constituencies (for example, database training for a particular college or user group)

In addition, some librarians track their reference data separately and include it in their self-evaluation during annual performance reviews as a measure of their contribution and productivity.

Though reference data are tracked and in many cases examined, comments from DLF respondents suggest that strategic planning is based on experience, anecdotes, and beliefs about future trends rather than on data. Several factors could account for this phenomenon. First, the data collected or compiled about reference service are, and will continue to be, incomplete. As one respondent observed, "Users ask anyone they see, so reference statistics will always be incomplete." Second, even if libraries have multiyear trend data on reference service, the data are difficult to interpret. Changes in institutional mission, the consolidation of reference points, the opening or renovation of library facilities, or the availability of competing "Ask-a" services could change either the use of reference service or its definition, service hours, or staffing. Decisions about what to count or not to count (for example, to begin including questions that go directly to librarians) make it difficult to compare statistics and interpret reference trends within an institution, let alone across institutions. Third, the technological environment blurs the distinction between reference, instruction, and outreach, which raises questions of what to count in which category and how to compile and interpret the data. Furthermore, libraries are creating frequently asked questions (FAQ) databases on the basis of their history of reference questions. What kind of service is this? Should usage statistics be categorized as reference or database use? Given the strenuous effort required to gather and compile reference data and the minimal use made of it, one wonders why so many libraries invest in the activity. One DLF site reported discontinuing gathering reference data based on a cost-benefit analysis.

1.2.5. Instruction

Librarians have traditionally offered instruction in how to use library resources. The instruction was provided in person—a librarian either visited a classroom or offered classes in the library. Often the instruction was discipline specific, for example, teaching students in a histo-

ry class how to use the relevant collections in the library. Digital library initiatives and the appearance of the Web have expanded both the content and format of library instruction. In addition to teaching users how to use traditional library resources, librarians now teach patrons how to use many different bibliographic and full-text electronic resources. Given concerns about undergraduate student use of the surface Web and the quality of materials they find there, library instruction has expanded to include teaching critical thinking and evaluation ("information literacy") skills. Remote access to the library has precipitated efforts to provide library instruction online as well as in person. The competencies required to provide instruction in the digital environment are significantly different from those required to teach users how to use traditional resources that have already been critically evaluated and selected by peer reviewers and librarians.

Libraries manually track their instruction efforts as a measure of another valuable service they provide to their constituencies. DLF respondents reported tracking the number of instruction sessions and the number of participants in these sessions. Sites with online courses or quizzes track the number of students who complete them. Libraries include instruction data in annual reports and national surveys. The data are used to monitor trends and to plan future library instruction. Some librarians track their instruction data separately and include this information in their self-evaluation during annual performance reviews as a measure of their contribution and productivity.

Though a substantial amount of work and national discussion is under way in the area of Web tutorials, national reporting mechanisms do not yet have a separate category for online instruction and no effort appears to have surfaced to measure the percentage of instruction offered online. Perhaps this is because the percentage is still too small to warrant measuring. Perhaps it is because online and in-person instruction are difficult to compare, since the online environment collapses session and participant data into one number.

1.2.6. Interlibrary Loan

Interlibrary loan (ILL) service provides access to resources not owned by the library. Libraries borrow materials from other libraries and loan materials to other libraries. The importance of ILL service to users and the expense of this service for libraries, many if not most of which absorb the costs rather than passing them on to users, lead to a great deal of data gathering and analysis about ILL. Changes precipitated by technology—for example, the ability to submit, track, and fill ILL requests electronically—expand data gathering and analysis.

Libraries routinely track the number of items loaned and borrowed, and the institutions to and from which they loan and borrow materials. They annually calculate the fill rate for ILL requests and the average turn-around time between when requests are submitted and the items are delivered. If items are received or sent electronical-

ly, the number of electronically filled requests (loaned or borrowed) and turn-around times are tracked separately. Some libraries also track the format of the items, distinguishing returnable items like books from non-returnable photocopies. Libraries that subscribe to OCLC Management Statistics receive detailed monthly reports of ILL transactions conducted through OCLC, including citations, whether requests were re-submitted, and turn-around times. They might have similar detail on ILL transactions conducted through other venues. Libraries with consortium resource-sharing arrangements track these transactions separately.

Some libraries track ILL requests for items in their own collections. Resource-sharing units that photocopy materials in their own collection and deliver them to campus users also track these transactions and, if a fee is charged, the revenue from these transactions. Libraries in multi-library systems track ILL activity at each library separately. If they operate a courier service among the libraries, they might also track these transactions.

Traditionally, much of this information has been tracked manually and later recorded in spreadsheets. The dual data entry is time-consuming and prone to human error. Implementing the ILLiad software enables automatic, detailed tracking of ILL transactions, saving staff time and providing a more complete and accurate picture of ILL activity.

ILL data are included in annual reports and national surveys. The data are used to

- Track usage and performance trends over time, including the percentage of ILL requests filled electronically
- Assess service quality on the basis of the success (fill) rate and average turn-around times
- Determine staffing on the basis of the volume of ILL or courier transactions throughout the year
- Distribute the ILL workload among libraries in a multilibrary system
- Inform requests for purchasing additional equipment to support electronic receipt and transmission of ILL items
- Target publicity to campus constituencies by informing liaison librarians about ILL requests for items in the local collection

One DLF respondent is considering analyzing data on ILL requests to assess whether requests in some academic disciplines are more difficult to fill than others are, though she did not explain how this data would be used. This respondent also wants to cross-correlate ILL data with acquisitions and circulation data to determine the number of items purchased on the basis of repeated ILL requests and whether these items circulated. Presumably this would enable a cost analysis of whether purchasing and circulating the items was less expensive than continuing to borrow them via ILL.

Cost data on ILL are important for copyright and budget reasons, but gathering the data to construct a complete picture of the cost of ILL transactions is complex and labor-intensive. Apparently many libraries have only a partial picture of the cost of ILL. Libraries

have to pay a fee if they borrow more than five articles from the same journal in a single year. Collecting the data to monitor this is difficult and time-consuming, and the data are often incomplete. Libraries that subscribe to OCLC Fee Management can download a monthly report of the cost of their ILL transactions through OCLC. Cost data for ILL transactions through other venues are tracked separately, and often not by the resource-sharing unit. For example, invoices for ILL transactions might be handled through the library's acquisitions unit; accounting for ILL transactions with institutions with which the libraries have deposit accounts might be handled through the administrative office. Often the cost data from these different sources are not compiled.

1.2.7. Printing and Photocopying

Printing and photocopying are important services provided by the library. Some libraries outsource these services, in which case they might not get statistics. If these services are under the library's control, they are closely monitored—particularly if the library does not recover costs. Printers and photocopies have counters that provide the number of pages printed or copied. The data are typically entered into a spreadsheet monthly. Some libraries also track the cost of paper and toner for printers and photocopiers. At least one DLF site even monitors the labor costs to put paper and toner in the machines. In some cases, use of these services by library staff and library users are tracked separately. The data are used to track usage trends and make projections about future use, equipment needs, expenditures, and revenue (cost recovery).

2. OUTCOME MEASURES

In the parlance of traditional library performance measures, the purpose of all inputs and outputs is to achieve *outcomes*. Outcomes are measures of the impact or effect that using library collections and services has on users. Good outcome measures are tied to specific library objectives and indicate whether these objectives have been achieved.[9] Outcomes assessments can indicate how well user needs are being met, the quality of library collections and services, the benefits or effectiveness of library expenditures, or whether the library is accomplishing its mission. Such assessments can be difficult and expensive to conduct. For example, how do you articulate, develop, and standardize performance measures to assess the library's impact on student learning and faculty research? Substantial work is underway in the area of outcomes assessment, but with the exception of ARL's LIBQUAL+, libraries currently have no standard definitions or instruments with which to make such assessments; likewise, they

[9] Bertot, J.C., C.R. McClure, and J. Ryan. 2000. *Statistics and Performance Measures for Public Library Network Services*. Chicago, Ill.: American Library Association: 66.

have no source of aggregate or contextual data to facilitate comparing and interpreting their performance. Given the difficulty and expense of measuring outcomes, if university administrators do not require outcomes assessments, many libraries do not pursue them.

2.1. Learning and Research Outcomes

No DLF respondent reported gathering, analyzing, or using learning and research outcomes data. Instead, they talked about the difficulty and politics of measuring such outcomes. Assessing learning and research outcomes is very difficult because libraries have no graduates to track (for example, no employment rate or income levels to monitor), no clear definitions of what to assess, and no methods to perform the assessments. The consensus among DLF respondents was that desirable outcomes or proficiencies aligned with the institutional mission and instruments to measure success should be developed through the collaboration of librarians and faculty, but the level of collaboration and commitment required to accomplish these two tasks does not exist.

In the absence of definitions and instruments for measuring learning and research outcomes, libraries are using assessments of user satisfaction and service quality as outcomes measurements. In the worst-case scenario, outputs appear to substitute for outcomes, but as one DLF respondent commented, "It's not enough to be able to demonstrate that students can *find* appropriate resources and are *satisfied* with library collections. Libraries need to pursue whether students are really *learning* using these resources." The only practical solution seems to be to target desired proficiencies for a particular purpose, identify a set of variables within that sphere that define impact or effectiveness, and develop a method to examine these variables. For example, conduct citation analysis of faculty publications to identify effective use of library resources.

2.2. Service Quality and User Satisfaction

Years ago, the Association of College and Research Libraries (ACRL) Task Force on Academic Library Outcomes Assessment called user satisfaction a "facile outcome" because it provides little if any insight into what contributes to user dissatisfaction.[10] Nevertheless, assessing user satisfaction remains the most popular library outcomes measurement because assessing satisfaction is easier than assessing quality. Assessments of user satisfaction capture the individual user's perception of library resources, the competence and demeanor of library staff, and the physical appearance and ambience of library facilities. In contrast, assessments of service quality measure the collective experience of many users and the gaps between their expecta-

[10] Task Force on Academic Library Outcomes Assessment Report. June 1998. Association of College and Research Libraries. p. 3. Available at: http://www.ala.org/acrl/outcome.html.

tions of excellent service and their perceptions of the service delivered. By identifying where gaps exist—in effect, quantifying quality—service quality studies provide sufficient insight into what users consider quality service for libraries to take steps to reduce the gaps and improve service. Repeating service quality assessments periodically over time can reveal trends and indicate whether steps taken to improve service have been successful. If the gaps between user perceptions of excellence and library service delivery are small, the results of service quality assessments could serve as best practices for libraries.

Though service quality instruments have been developed and published for several library services, the measure has had limited penetration. Few DLF sites reported conducting service quality assessments of particular services, though many are participating in ARL's LIBQUAL+ assessment of librarywide service provision. DLF libraries reported conducting service quality studies of reference, interlibrary loan, course reserves, and document delivery services to assess user perceptions of their speed, accuracy, usefulness, reliability, and courteousness. The results were used to plan service improvements based on identified gaps. In some cases, the results were not systematically analyzed—additional examples of a breakdown in the research process that leads to purposeless data collection. One DLF respondent suggested that the best approach to measuring service quality using the gap model is to select which service to evaluate on the basis of a genuine commitment to improve service in that area, and then define quality in that area in a way that can be measured (for example, a two-day turn-around time). The keys are commitment and a clearly articulated measurable outcome.

DLF respondents raised thought-provoking philosophical questions about assessments of service quality:

- Should service quality assessments strictly be used as diagnostic tools to identify gaps, or should they also be used as tools for normative comparison across institutions?
- Do service quality assessments, designed to evaluate human-to-human transactions, apply to human-computer interactions in the digital environment? If so, how?
- Are human expectations or perceptions of quality based on facts, marketing, or problems encountered? How do libraries discover the answer to this question, and what are the implications of the answer?
- If quality is a measure of exceeding user expectations, is it ethical to manage user expectations to be low, then exceed them?

2.3. Cost-Effectiveness and Cost Benefits

Libraries have traditionally tracked costs in broad categories, for example, salaries, materials, or operating costs. ARL's E-metrics initiative creates new categories for costs of e-journals, e-reference works, e-books, bibliographic utilities, and networks and consortia, and even of the costs of constructing and managing local digital collec-

tions. Measuring the effectiveness and benefits of these costs or expenditures, however, is somewhat elusive.

"Cost-effectiveness" is a quantitative measure of the library's ability to deliver user-centered outputs and outcomes efficiently. Comments from DLF respondents suggest that the only motivation for analyzing the cost-effectiveness of library operations comes from university administrators, which is striking, given the budgetary concerns expressed by many of the respondents. Some libraries reported no impetus from university administrators to demonstrate their cost-effectiveness. Others are charged with demonstrating that they operate cost-effectively. The scope of library operations to be assessed and the range of data to be gathered to assess any single operation are daunting. Defining the boundaries of what costs to include and determining how to calculate them are difficult. Published studies that try to calculate the total cost of a library operation reveal the complexity of the task and substantial investment of time and talent required to assemble and analyze a dizzying array of costs for materials, staffing, staff training, hardware, software, networking, and system maintenance.[11] Libraries charged with demonstrating their cost-effectiveness are struggling to figure out what to measure (where to begin), and how to conduct these assessments in a cost-effective manner.

Even if all of the costs of different library operations can be assembled, how are libraries to know whether the total cost indicates efficient delivery of user-centered outputs and outcomes? In the absence of standards, guidelines, or benchmarks for assessing cost-effectiveness, and in many cases a lack of motivation from university administrators, an ad hoc approach to assessing costs—rather than cost-effectiveness—is under way. DLF respondents reported practices such as the following:

- Analyzing the cost per session of e-resource use
- Determining cost per use of traditional materials (based on circulation and in-house usage statistics)
- Examining what it costs to staff library services areas
- Examining what it costs to collect and analyze data
- Examining the cost of productivity (for example, what it costs to put a book on the shelf or some information on the Web)
- Examining the total cost of selected library operations

The goals of these attempts to assess costs appear to be to establish baseline data and define what it means to be cost-effective. For example, comparing the cost per session of different e-resources can facilitate an understanding of what a cost-effective e-resource is and perhaps enable libraries to judge vendor-pricing levels.

Cost-benefit analysis is a different task entirely because it takes into account the qualitative value of library collections and services

[11] See, for example, C.H. Montgomery and J. Sparks, Framework for Assessing the Impact of an Electronic Journal Collection on Library Costs and Staffing Patterns. Available at: http://www.si.umich.edu/PEAK-2000/montgomery.pdf.

to users. Even if libraries had a clear definition of what it means to be cost-effective or a benchmark against which to measure their cost-effectiveness, additional work is required to determine whether the benefits of an activity warrant the costs. If the cost of an activity is high and the payback is low, the activity may be revised or abandoned. For example, one DLF respondent explained that his library stopped collecting reference statistics in 1993, when it determined that the data seldom changed and it cost 40 hours of staff time per month to collect. Quantifying the payback is not always so straightforward, however. User studies are required to assess the value to users of seldom-used services and collections. Knowing the value may only raise the question of how high the value must be, and to how many users, to offset what level of costs.

The terrain for conducting cost-benefit analyses is just as broad and daunting as is the terrain for assessing cost-effectiveness of library operations. One DLF institution is analyzing the costs and benefits of staff turnover, examining the trade-offs between loss of productivity and the gains in salary savings to fund special projects or pursue the opportunity to create new positions. As with analyses of cost-effectiveness, libraries need guidelines for conducting cost-benefit analyses and benchmarks for making decisions. The effort requires some campuswide consensus about what it values about library services and what is worth paying for.

BEST COPY AVAILABLE